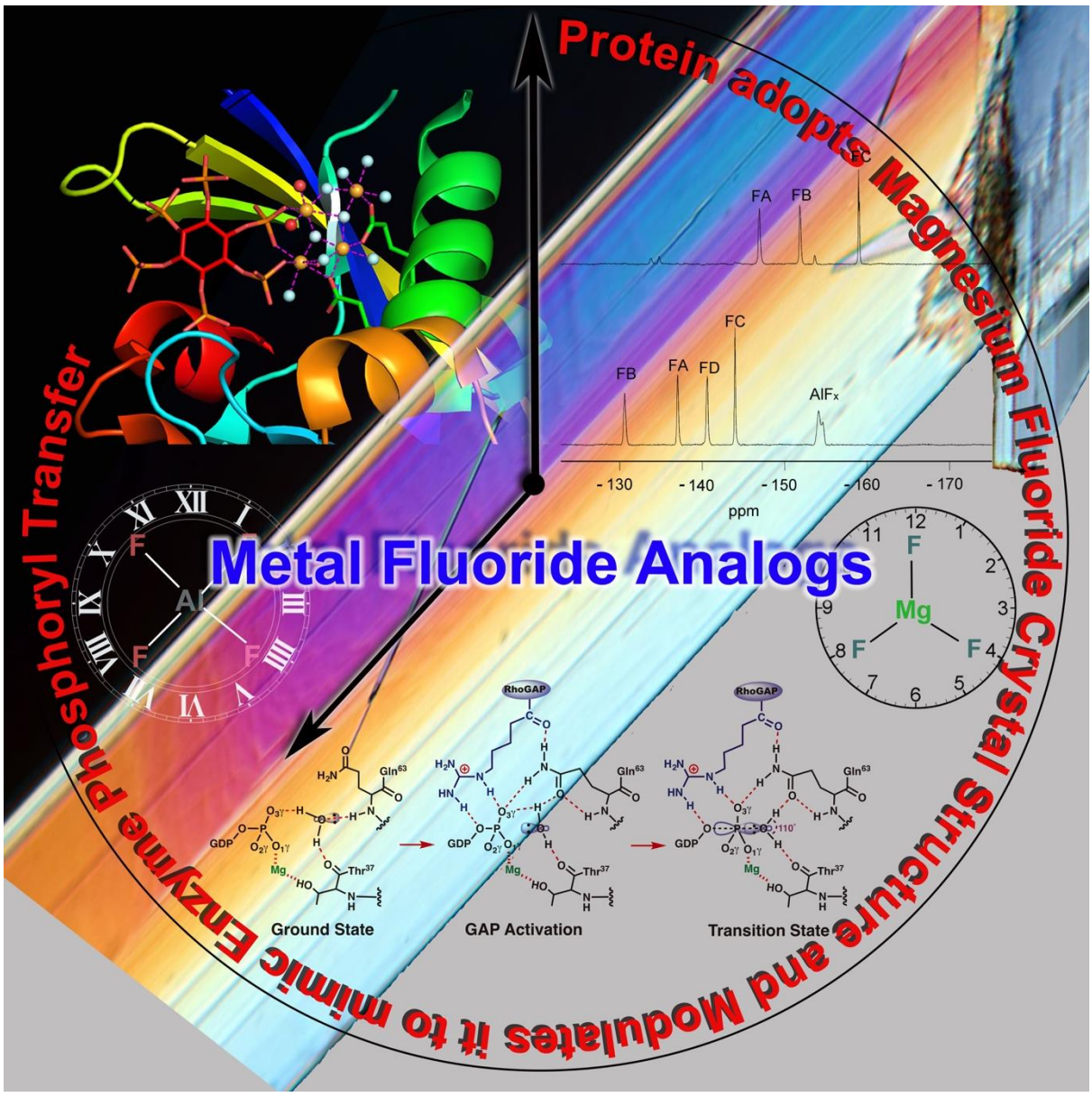
Metal Fluorides as Analogs for Studies on Phosphoryl Transfer Enzymes

Yi Jin[a], Nigel G. Richards[b], Jonathan P. Waltho[c], and G. Michael Blackburn[d]*.



Abstract: In 1994, the protein structure of a transition state analogue for G1α, a small G protein, heralded a new field of research into the structure and mechanism of enzymes that manipulate transfer of the phosphoryl (PO_3^-) group. It was based on a protein complex of GDP and AlF_4^- that mimicked the transition state for hydrolysis of GTP. The growing list of enzyme structures that embrace metal fluorides, MF_x, as ligands that imitate either the phosphoryl group or a phosphate, now exceeds 80 per triennium. They fall into three distinct geometrical classes: (i) Tetrahedral complexes based on BeF_3^-, mimic ground state phosphates, (ii) Octahedral complexes, primarily AlF_4^-, mimic "in-line" anionic transition state for phosphoryl transfer, and (iii) Trigonal bipyramidal complexes additionally mimic the tbp stereochemistry of the transition state and are represented by MgF_3^- and putative AlF_3^0 complexes. Their interpretation has provided a deeper mechanistic understanding of the behavior and role of phosphate monoesters in molecular biology. This review challenges the existence of AlF_3^0 and $MgF_4^=$ as real species in protein complexes and questions the relevance for enzymes of physical organic chemistry and model studies that are water-based for phosphoryl group transfer. It proposes a new interpretation of the role of general acid-base catalysis.

1. Introduction

For a decade following their discovery, the atomic structures of proteins containing a metal fluoride (MF_x) species were based primarily on geometric considerations. From 2003 onwards, this resulted in a growing uncertainty about their chemical constitution. Now, ^{19}F NMR analysis of these complexes has been used firstly to analyze and identify their atomic composition, secondly to establish their significance in solution, thirdly to deliver experimental measurements of the electronic environment provided by the protein in conformations close to the transition state (TS), and lastly to identify a significant number of mis-assignments, thereby providing a corrective critique for past errors and future uncertainties.

There are now over 500 MF_x structures in the Protein Data Bank (PDB) (Fig. 1). The molecular analysis of these structures has established a simple, logical, and rational understanding of the chemical constitution of transition state analog (TSA) and ground state analog (GSA) structures of MF_x complexes. It provides a firm base for understanding enzymatic mechanisms for the catalysis of phosphate monoesters and anhydrides, notably ATPases, GTPases, kinases, mutases, phosphatases, and phosphohydrolases.[1] They all employ "in-line" geometry, they are concerted, and they utilize tight control of hydrogen bonding in the active site complex to disfavor the formation of hydrogen bonds that would inhibit the chemical step in catalysis. Some of

[a] Dr. Y. Jin , Department of Chemistry,
University of York
York, YO10 5DD(UK)
[b] Prof. Dr. N.G. J. Richards
School of Chemistry
Cardiff University
Cardiff, CF10 3AT (UK)
[c] Prof. Dr. J. P. Waltho
Manchester Institute of Biotechnology
Manchester, M1 7DN (UK)
[d] Prof. Dr. J. P. Waltho, Prof. Dr. G. M. Blackburn
Department of Molecular Biology and Biotechnology,
University of Sheffield
Sheffield, S10 2TX (UK)
email: g.m.blackburn@sheffield.ac.uk

Yi Jin took her BSc chemistry in Xiamen University, China followed by an MSc in organic chemistry in the group of Prof. Yufen Zhao. She received her PhD in 2012 followed by a one-year postdoc under the guidance of Profs. J.P. Waltho and G. M. Blackburn in the University of Sheffield, UK. In 2014, she moved to the group of Prof. G. J. Davies in YSBL, University of York, UK. Her research interests involve mechanistic studies of disease-relevant phosphoryl transfer enzymes and carbohydrate processing enzymes using chemical, NMR, molecular biology, and crystallographic approaches.

Michael Blackburn is Emeritus Professor of Biomolecular Chemistry in Sheffield University and a founding member of the Krebs Institute. His undergraduate and post-doctoral career in Cambridge University led him in 1961 into the biological chemistry of phosphorus under Alexander Todd. He has worked on nucleotides, their analogs, and the enzymes that use them at the interface of chemistry and molecular biology for over 50 years. Most recently, he has focused on the unique paradox between their structural stability and kinetic lability.

Nigel Richards is Professor of Biological Chemistry at Cardiff University and a Research Fellow at the Foundation for Applied Molecular Evolution in Gainesville, Florida. He received his undergraduate and doctoral degrees from Imperial College, London and Cambridge University, respectively. After post-doctoral studies at Columbia University, where he was an author of the MacroModel software package, he held academic positions in the UK and USA. His current research interests are the design of enzyme inhibitors to investigate cellular metabolism in sarcomas, and the elucidation of catalytic mechanisms employed by Fe- and Mn-dependent enzymes.

Jonathan Waltho is the Gibson Professor of Biophysics at the University of Sheffield, and Professor of Structural Biology at the University of Manchester. He received his undergraduate degree from Durham University and his doctoral degree from Cambridge University. Following post-doctoral studies at SmithKline French and The Scripps Research Institute, he has tackled a wide range of problems in protein folding, misfolding, and protein-ligand interactions. His current research interests focus on structure, electronics and dynamics in phosphoryl, methyl and hydride transfer enzymes.

these structures have been starting points for multiple studies on enzyme mechanisms using QM-MM and DFT analysis. More controversially, the present analysis challenges the extrapolation to enzymes of physical organic chemistry, of model studies, and the role of general acid-base catalysis (GABC) of phosphoryl transfer in aqueous solution. structures, well supported by ^{19}F NMR data and the best of many computational developments.

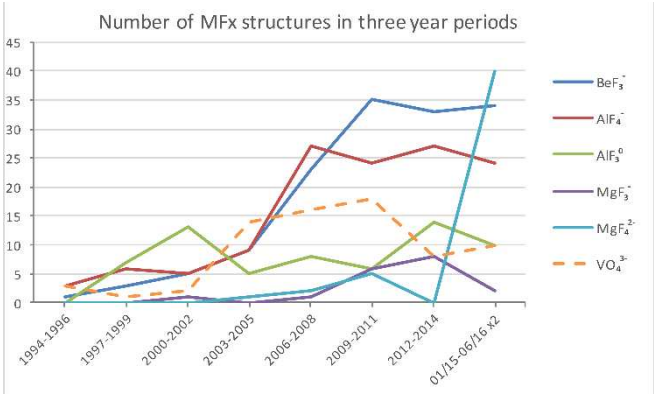
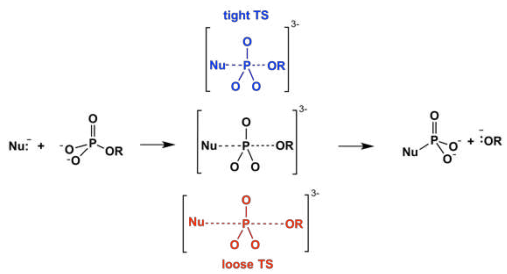


Figure 1. Number of MF_x structures published in the PDB triennially, vanadate data included for reference (data for 01/15 through 06/16 normalized by x2 to represent a triennial figure).

Phosphoric acid (H_3PO_4), its esters, amidates, and anhydrides share a common tetrahedral geometry based on a phosphorus (V) core linked near-symmetrically to four oxygens or nitrogens. Biological phosphoryl transfer reactions call for the relocation of a phosphoryl group, PO_3^-, from a donor to an acceptor atom, typically N, O, or S and more rarely C or F. There are many reviews of this activity and its catalysis, but there is no consensus on whether the reactions are more associative (tight TS) or more dissociative (loose TS) in character, a description with a boundary value of 4.9 Å based on van der Waals considerations (Scheme 1).[1a] In either case, the TS for such phosphoryl group transfer will have trigonal bipyramidal (tbp) geometry, with axial dimensions that relate to the tight or loose nature of the TS. Because the primary database for MF_x complexes is structurally driven, we review the separate groups of MF_x protein complexes in terms of their geometry. This has the additional advantage of overriding ambiguities in the assignment of composition, as shown below.



Scheme 1. Concerted phosphoryl transfer. Top, bond making precedes bond breaking (blue); center, bond breaking balanced by bond making (black); bottom, bond breaking in advance of bond making (red).

2. Tetrahedral Phosphate Mimics, BeF_3^-

Beryllium (II) forms stable fluorides in water that exist as a mixture of tetrahedral species including $BeF_2{\cdot}2H_2O$, $BeF_3^-{\cdot}H_2O$, and $BeF_4^=$.[2] Early NMR studies on fluoroberyllate complexes with ADP led to analysis of mixed fluoroberyllate·ADP species with myosin and the first x-ray analysis of a fluoroberyllate protein structure was delivered in 1995 for an ADP·BeF_3^- complex with myosin (PDB: **1mmd**).[3] Since then, 122 trifluoroberyllate complexes have been described, with 3 solved by NMR and 119 x-ray structures having resolutions ≥ 1.2 Å. The vast majority of these structures have a tetrahedral trifluoroberyllate bonded to anionic oxygen. They divide into two principal groups: over 70 are coordinated to an aspartate carboxylate (including the 3 NMR structures) and around 50 are coordinated to a nucleotide terminal phosphate. Only 2 are coordinated to a histidine ring nitrogen.

2.1. Aspartyl trifluoroberyllates

These structures share a common core, with bidentate coordination to an essential metal ion, generally Mg^{2+} and rarely Mn^{2+}, from fluorine F1 and the second carboxylate oxygen, OD2, giving a near planar six-membered ring (Fig. 2). [Here, and throughout, naming of atoms in phosphates and their analogues conforms to IUPAC 2016 recommendations].[4] Beryllium is difficult to locate by x-ray diffraction because it has low electron density. This results in uncertainty in its location, resulting in considerable variation in attributed geometry (Fig. 2). Pauling assigned predominantly ionic character to the Be–F bond (80%), leading to expectation of solvation of the trifluoroberyllate function by water.[5] However, only 10 of the 30 best resolved structures show such an isolated water proximate to the BeF_3^- moiety, which is not “in line” with the O–Be bond (155.3 ± 9.2˚), and is at widely variable distance from the beryllium atom (3.8 ± 0.5 Å) (Fig. 2, SI Table. 1).

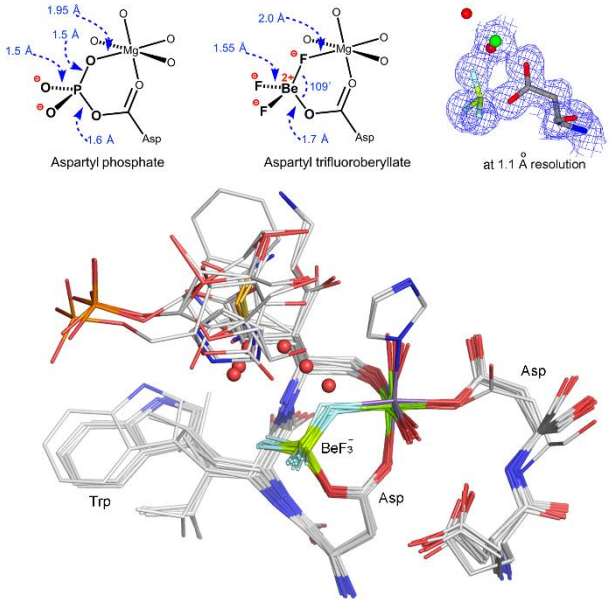


Figure 2. 17 Aligned aspartyl-trifluoroberyllate structures. BeF_3^- is locked in a 6-membered ring (center). Catalytic Mg^{2+} (rarely Mn^{2+}) and an aspartate (usually Asp57) fuse a 13-membered ring to the fluoroberyllate ring with atoms from the adjacent two amino acids downstream (rear center). Octahedral coordination to Mg is completed by an additional aspartate (right), by 1-2 waters, but only twice by histidine (upper right). (Atom colors: fluorine, light blue; beryllium, yellow-green; nitrogen, blue, oxygen, red). In 7 structures, an isolated water (red spheres) is distantly related to one fluorine. (Electron densities presented in CCP4MG from mtz data in EDS and contoured at 1σ)

2.2. ADP•BeF_3^- structures

There are 42 x-ray structures of BeF_3^- complexes with ADP and 6 with GDP, which constitute isosteric mimics of ATP and GTP respectively. They are distributed among kinases, hydrolases, mutases, helicases, and small G proteins. Of the ADP·BeF_3^- structures, 25 are resolved at ≤ 2.5 Å and 20 align remarkably well (Fig. 3). The beryllium is bonded to O3B and a catalytic Mg^{2+} is coordinated to F1 and to O1B in a 6-membered ring. There is remarkable consistency in neighboring amino acids; an arginine and a lysine coordinate β and γ and balance the anionic charge of the nucleotide. By contrast, the adenine base occupies a range of conformations (Fig. 3, SI Table 2). A very significant feature is that 12 of the 20 structures have a water H-bonding to one of the three fluorines. These waters lie well within the BeF_3^- "cone" with their oxygen being ~3.4 Å from the beryllium, with a median "in-line" angle of 158°, and forming a H-bond to one of the fluorines (2.82 ± 0.27 Å). As the axial O—Be—O distance is close to 5.1 Å, these waters are part of a Near Attack Conformation (NAC) that is intermediate between a ground state (GS) and a TS situation.[6]

The 6 GDP structures are very similar in structure to the ADP complexes but at rather lower resolution (SI Table 3).

The BeF_3^- complex for human phosphoglycerate kinase (hPGK) raises the question: "Where is the beryllium in the case of two oxyanion acceptors?" The structure of the complex hPGK·ADP·BeF_3^-·3PG (PDB: **4axx**, 1.74 Å resolution) places the Be atom 1.73 Å from the carboxylate oxygen and 2.85 Å from the ADP oxygen O3B. However, the three fluorines are on average 2.75 Å from the carboxylate oxygen and 2.96 Å from the ADP oxygen (Fig. 4A). As the sum of van der Waals radii for Be–O is 3.26 Å, these data suggest mixed occupancy with beryllium closer on average to the carboxylate.[7]

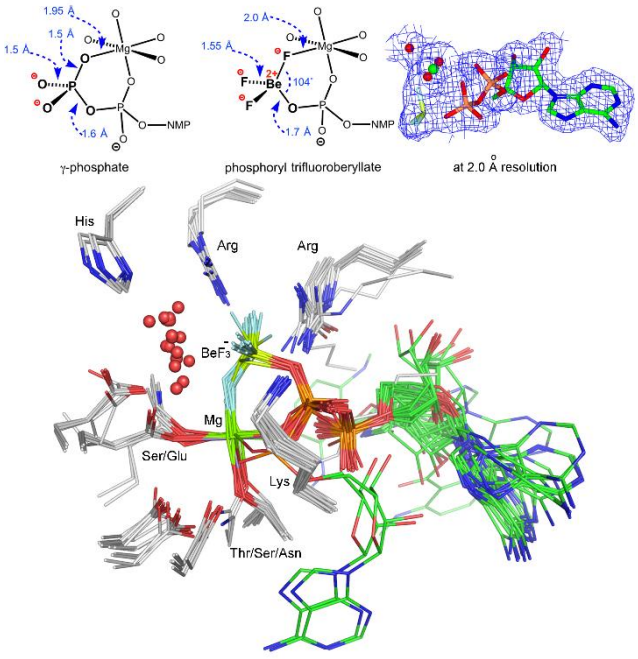


Figure 3. In 20 aligned ADP·trifluoroberyllate structures, BeF_3^- is locked in a 6-membered ring (center) with catalytic Mg^{2+} coordinating F1 and O3B. Octahedral coordination to Mg is completed to OB1, by 2 *trans*-waters (not shown), by a Ser/Glu side chain oxygen and a Ser/Thr/Asn side chain oxygen. γ-Phosphate coordination to an Arg and a Lys is also common. Location of adenines is very variable (in green). In 12 structures, an isolated water (red spheres) is located close to the BeF_3^- "cone". Atom colors: fluorine, light blue; beryllium, yellow-green; nitrogen, blue; oxygen, red). The protein residues are in gray. **NB** There is *prima facie* evidence that two of these (PDB: **1w0j** and **4znl**) may really be trifluoromagnesate structures. This is because (a) their geometry is "in-line" and the O—M—O distance is short and (b) the crystallization mix contained ≥ 100 mM of a beryllium sequestering component: citrate or EDTA).

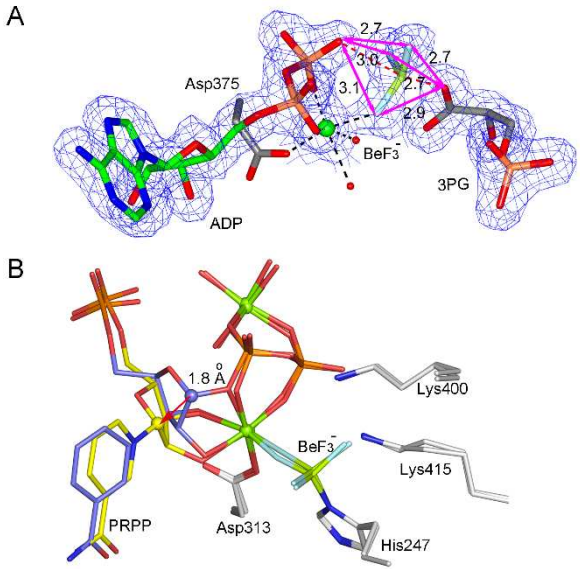


Figure 4. (A) Structure of BeF_3^- complex for hPGK (PDB:4axx). Beryllium (lime stick) is "in-line" between ADP and 3PG. The non-bonding fluorine-to-oxygen distances (magenta arrows) are shorter to the carboxylate than to the ADP oxygen. (B) Nicotinamide phosphoribosyltransferase catalyses displacement of pyrophosphate from C1 of ribose 5-phosphate (reactants in purple, products in silver, red curly arrow shows departure of phosphoryl oxygen). Structures of 2 overlaid complexes show BeF_3^- bonded to Nε of His247 and one fluorine coordinating octahedral Mg^{2+} (green sphere). C1' of PRPP in reactant (purple sphere) moves 1.8 Å to bond the nicotinamide N1 (silver sphere), Reactant purple sticks, product silver sticks, Be in yellow-green.

2.3 Histidine trifluoroberyllates

Various approaches to analogs of τ–phosphohistidine have been explored. Work on nicotinamide phosphoribosyltransferase (NAMPT) has structurally mimicked phosphorylation of an active-site histidine using trifluoroberyllate. Crystal structures of NAMPT for reactant and product complexes (Fig. 4B) have a covalent His247·BeF_3^-. In contrast to all other trifluoroberyllate structures, it is coordinated to one fluorine without any direct linkage to His247.[8]

2.4 A nucleotide beryllium difluoride structure

A solitary example of beryllium difluoride bridging ADP and UDP illuminates the activity of UMP/CMP kinase.[9] The 2.0 Å structure (Fig. 5A) has a tetrahedral beryllium bridging O3B of ADP to O1B of UDP. An essential Mg^{2+} coordinates one fluorine, and O1B of ADP. The two diastereotopic fluorines show well-separated resonances in the ^{19}F NMR (Fig. 2B). This stable mimic of Ap_5U is strongly coordinated to 4 arginines and 1 lysine and thus endorses the observation that nucleotide kinases are more strongly inhibited by Ap_5Nuc than by Ap_4Nuc on account of their additional negative charge.[10]

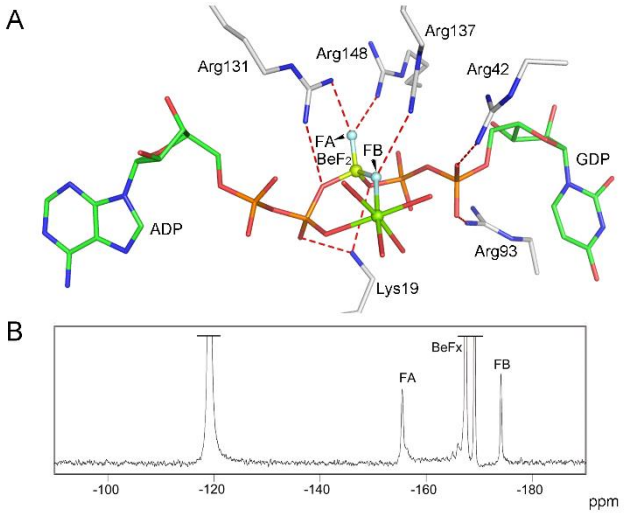


Figure. 5 (A) Structure of BeF_2 complexed to 2 nucleotides in U/C kinase. Beryllium (olive sphere) is bonded to oxygens of ADP (green) and UDP (purple) with one fluorine (light blue) coordinating an octahedral Mg^{2+} (green sphere). The tetrahedral complex is coordinated by 5 H-bonds to 4 amino acids. (gray sticks). (B) ^{19}F NMR for the ADP·BeF_2·UDP complex as above.

2.5 Conclusions

The significant ability of beryllium (II) fluorides to complete tetrahedral coordination by binding to an anionic oxygen makes them effective isosteric and electrostatic analogs of phosphate in a wide range of situations.[11] The bond lengths for Be–F and Be–O are close to those for P–O (1.6 ± 0.5 Å) and the dominant ionic character of the Be–F bond[5] means that the fluorines readily accept H-bonds from a range of donors and/or coordinate to Group 2 metal ions. These mimics have been advantageously used to study changes in major conformation of proteins by crystallography, ^1H NMR, and EM while studies on ADP·BeF_3^- have supported investigations of ATPases that drive various mechanical processes at a molecular level, particularly for myosin.[12] They have proved especially valuable for the identification of near attack conformations (NACs) in enzyme mechanisms, especially for βPGM.[13]

3. Octahedral MF_x Complexes

Aluminum (III) forms stable fluorides in water that exist as a mixture of octahedral species including $AlF_2^+·4H_2O$, $AlF_3·3H_2O$, $AlF_4^-·2H_2O$, and $AlF_5^{=}·H_2O$ depending on the concentration of fluoride.[14] Their stability is a function of pH because aluminum forms insoluble $Al(OH)_3$ above pH 7.5.[14] Aluminum and fluoride were were discovered to stimulate the activity of small G proteins in the presence of GDP,[15] and the proposal that they could mimic the active GTP bound state[16] was endorsed by ^{19}F NMR analysis, which identified the formation of a GDP·AlF_x complex for $G_1\alpha$.[17] In 1994, crystal structures for tetrahedral GDP·AlF_4^- complexes of transducin α and a small G protein, $G_i\alpha_1$, appeared almost simultaneously, and were soon followed by an ADP·AlF_4^- structure for a myosin fragment.[3a, 18] Since then, the number of such AlF_4^- complex structures in the PDB (PDB ligand: **ALF**) determined by crystallography has grown steadily to reach 109 by March 2016 (Fig. 1, SI Table 4).

3.1.1 Aspartyl tetrafluoroaluminates

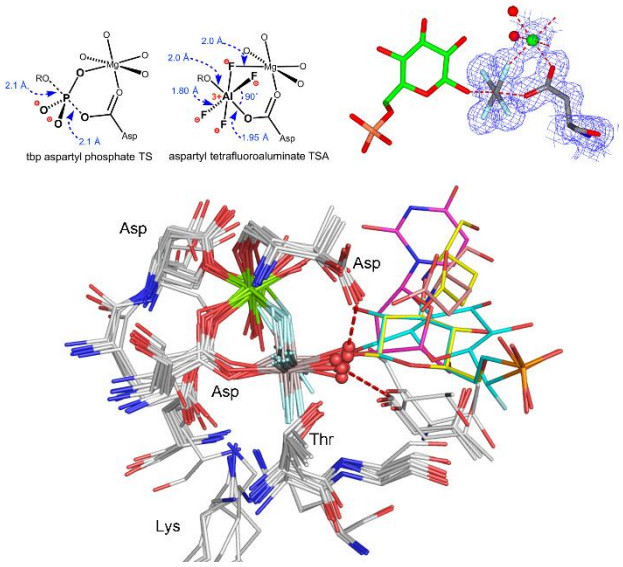


Figure. 6. Structures of 14 aspartyl tetrafluoroaluminates superposed by Cα alignment. Aluminum is octahedrally coordinated to Asp-O4 (gray) forming a 6-membered ring with a catalytic magnesium and "in-line" with the acceptor oxygen, water (red sphere) or the hydroxyl group of a nucleoside or hexose reactant (colors). (Atom colors: fluorine, light blue; aluminum, gray; nitrogen, blue; oxygen, red; magnesium, green).

The PDB has 14 structures with a tetrafluoroaluminate bonded to an aspartyl oxygen. This mimics an aspartyl phosphate, known to be a transient species in the catalytic activity of these enzymes. They have an essential Mg^{2+} enclosed in a 6-membered ring, as seen for the corresponding BeF_3^- structure (Section. 2.2), and all align very well on PDB: **2wf7** (Fig. 7, SI Table 6), showing commonality of remaining 4 ligands coordination the catalytic Mg^{2+}. These structures fall into two subsets: six members of the first group have a second aspartate next-but-one to the first, and it coordinates the oxygen that is the sixth aluminum ligand. The O–Al–O bonds are "in-line" (167.5° ± 7.0°) with the aluminum midway between the two oxygens (separation 3.94 ± 0.11 Å). The Al-F bonds are 1.78 ± 0.02 Å (for the 6 best-resolved structures), independent of coordination to Mg. βPGM accounts for three of the six structures, the other three being a human mitochondrial deoxyribonucleotidase, a phosphoserine phosphatase (PSP), and a C-terminal domain phosphatase that operates on RNA polymerase II. In all of these, a catalytic aspartate accepts a short H-bond from the apical water/hydroxyl group (2.59 ± 0.05 Å) to complete the orientation of this oxygen for nucleophilic attack on the aspartyl phosphate.[19]

The second subset comprises ATPases involved in pumping Ca, Cu, and Zn. They use an aspartyl phosphate intermediate, whose TS for hydrolysis is mimicked by the octahedral AlF_4^-. These have "in-line" O–Al–O bonds (163.8° ± 8.1°) with aluminum midway between the two oxygens (O-O separation 3.92 ± 0.14 Å) and Al-F bonds 1.78 ± 0.02 Å. An axial water oxygen forms short H-bonds to an invariant glutamate (2.54 ± 0.1 Å) and to a threonine carbonyl (2.57 ± 0.05 Å). These residues clearly orientate and

polarize the water for "in-line" attack on the aspartyl phosphate (Section 8.3).[20]

3.1.2. Nucleotide tetrafluoroaluminates, GDP

There are 46 x-ray structures of AlF_4^- complexes with GDP that constitute isoelectronic but non-isosteric mimics of GTP in small G proteins, dynamins, ribosomal factors, kinases, ATPases, mutases, ion pumps, and helicases. Of these structures, 25 are resolved at ≤ 2.7 Å and align remarkably well (Fig. 7, SI Table 5). The aluminum is bonded to GDP by O3B and the catalytic Mg^{2+} is coordinated to F1 and O1B in a 6-membered ring. There is remarkable consistency in neighboring amino acids, notably by a heptapeptide near the N-terminus, sequence XXXXGKS(T), whose serine hydroxyl coordinates magnesium *trans* to a fluorine. The guanosine base and ribose occupy a common conformation (Fig. 7) with the exception of Atlastin (PDB: **4ido**). The geometry of the AlF_4 moiety is well defined, being regularly octahedral to 2.7 Å resolution, with an "in-line" O–Al–O angle 172.8° ± 7.1°, with aluminum midway between the axial oxygens that are 4.07 ± 0.23 Å apart (Table 1), and have AlF bonds 1.77 ± 0.28 Å. All the structures have an axial oxygen ligand (Fig. 7, red spheres) to aluminum that is trigonal planar with respect to two H-bond acceptors (ψ–dihedral 4.9° ± 2.9°) whose angle to the axial oxygen is 102 ± 6°(Fig. 8). One is the backbone carbonyl of a threonine, whose OG coordinates the magnesium (Fig. 7, upper right). The second is a glutamine side-chain carbonyl or a water (Fig. 7, lower right, red spheres).

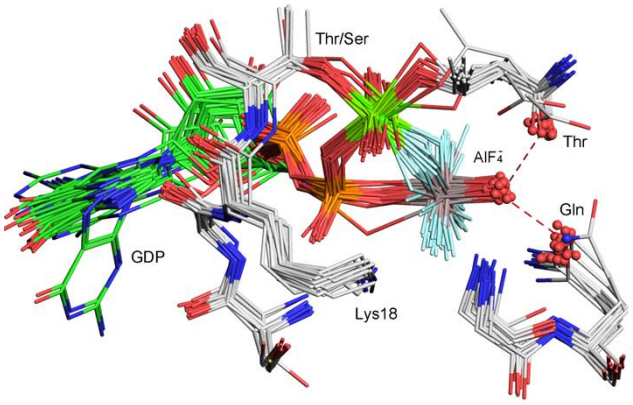


Figure. 7 GDP tetrafluoroaluminate structures. 25 Structures are superposed on (PDB: **2gj8**) by α-carbon atoms (primarily for the invariant heptapeptide, bottom to top center). AlF_4^- is locked in a 6-membered ring (center) with catalytic Mg^{2+} coordinating F1 and O3B. Octahedral coordination to Mg^{2+} is provided by OB1, 2 *trans*-waters, a Thr hydroxyl (top right), and a Ser/Thr hydroxyl (top center). Phosphate oxygen coordination to a Lys18 (center) is standard. Location of guanines is regular (left rear) with two exceptions. Atom colors: fluorine, light blue; aluminum, gray; nitrogen, blue; oxygen, red; magnesium, green).

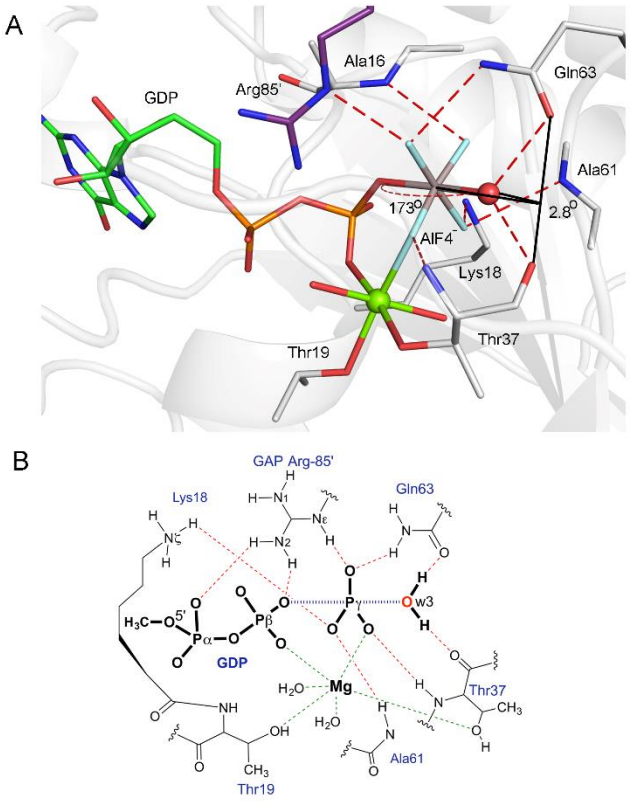


Figure. 8 (A) RhoA/RhoGAP·GDP·AlF_4^- complex (PDB: **1tx4**) showing H-bonding from nucleophilic water to carbonyl oxygens of Gln63 and Thr37 with a ψ-dihedral angle 2.8° and in-line angle 173.0°. (Atom color: carbon, silver, aluminum, gray; nitrogen, blue; oxygen, red; fluorine, sky blue; magnesium, green). (B) Scheme to show H-bond network for RhoA/GAP·GTP·wat TS complex.

3.1.3. Nucleotide tetrafluoroaluminates, ADP

The 45 octahedral structures that have AlF_4^- bonded to a terminal oxygen of ADP (O3B) include kinases, hydrolases, isomerases, myosins, helicases, transporter pumps, and nitrogenase. They mimic ATP and are relatively diverse in conformation. The 25 that are resolved at ≤ 2.5 Å have an axial O–Al–O distance of 4.04 ± 0.14 Å with an "in-line" angle of 170° ± 8°. The majority of the 45 have a water as the second oxygen ligand with the catalytic Mg^{2+} also coordinated to one β-oxygen and a fluorine. This is illustrated for F1ATPase (PDB: **1h8e**) (Fig. 9A). Three complexes have magnesium triply coordinated to OA, OB and F.

Overall, the aluminum is a little closer to O3B (1.97 ± 0.12 Å) than to the second oxygen (2.10 ± 0.12 Å), and Al-F bond lengths (for the 12 best-resolved structures) are 1.77 ± 0.04 Å. The variable general position of the fluorines relative to the catalytic Mg^{2+} suggests that some compromise has been reached in fitting four fluorines into protein loci that have evolved to accommodate three electronegative oxygens.

3.1.4. Other tetrafluoroaluminates

Two structures have AlF_4^- bonded to a histidine nitrogen, as illustrated for phosphoglycerate mutase (PDB: **2f90**). This mimics phosphoryl transfer from His11 to OH-2 of 3PGA (Fig. 9B).

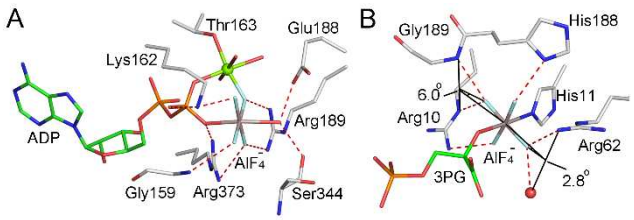


Figure. 9 (A) F1ATPase TSA complex with ADP·AlF_4^-·Wat showing local charge balance for 5 ±ve and 5 –ve charges. (B) AlF_4^- TSA complex mimicking phosphoryl transfer from His11 to 3PGA OH-2. Aluminum coordinated four fluorines with His11 Nε and PGA OH-2 as axial ligands (Atom colors: 3PGA, green; fluorine, pale blue; amino acids, silver).

3.2 Trifluoroaluminates, AlF_3^0

There are three examples of octahedral complexes where an aluminum trifluoride core is expanded to octahedral, six-coordination by having three oxygen ligands (SI Table 7). For the small G protein Rab5a, the mutation A30P results in the addition of the side chain hydroxyl of Ser29 to aluminum. For hPGK, the mutation K219A results in the addition of water to the aluminium. For a bacterial dUTPase, aluminum trifluoride takes the place of the β-phosphoryl group in dUTP and coordination to O3A, O3B, and to the water nucleophile completed the octahedral array (Fig.10). This structure provides a unique example where nucleophilic attack is not directed at a terminal phosphorus.[21]

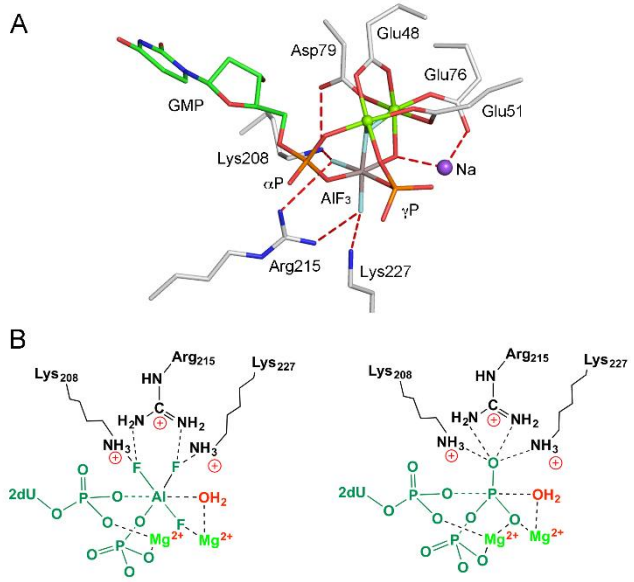


Figure. 10 (A) Trifluoroaluminate structure for dUTPase (PDB: **4di8**). AlF_3 coordinates GMP (green bonds) within-line water co-ordinated to sodium (purple sphere) and with $PO_4^=$ adjacent to the leaving O3A. Two magnesiums (green spheres) are located by coordination to the reactants and to four carboxylate residues. (B) Cartoon showing octahedral AlF_3 sharing the tbp coordination of the true TS for a phosphoryl group.

4. Trigonal Bipyramidal MF_x.

4.1. Trifluoromagnesates MgF_3^-

Magnesium does not form multiple stable fluorides in water. They are moderately soluble (2 mM) with a dissociation constant for $MgF_{2(aq)}$ estimated at 10^{-5} M.[22] However, trifluoromagnesate protein complexes were first anticipated on the basis of magnesium-dependent fluoride inhibition studies that in 2002 led to the first identification of MgF_3^- in a tbp crystalline TSA complex for the small G protein RhoA/RhoGAP (Fig. 11A).[23] The PDB now lists 16 entries for this ligand (code MGF) while a further 3 entries assigned as AlF_3^0 have been shown by ^{19}F NMR to be MgF_3^- complexes (SI Table 8).[24] Magnesium is regularly 6-coordinate and gives octahedral complexes with oxygen ligands. By contrast, trifluoromagnesate is 5-coordinate, trigonal bipyramidal in proteins, and has ideal characteristics to mimic the phosphoryl group as it is isoelectronic with PO_3^- and has similar geometry. Examples of its use include small and large molecule kinases, mutases, phosphatases, and hydrolases. Their complexes invariably involve coordination to one catalytic Mg^{2+} (two for some protein kinases), which are usually in a cyclic 6-membered ring structure, as shown for aspartyl phosphate mimics (Fig.11B). They have an axial O-Mg-O distance of 4.19 ± 0.08 Å with an in-line angle 171.4° ± 3.9°. The axial Mg–O bonds are 2.13 ± 0.10 Å with Mg-F bonds of 1.83 ± 0.06 Å, compared to computed non-bridging P–O bonds of 1.52 ± 0.02 Å.[25]

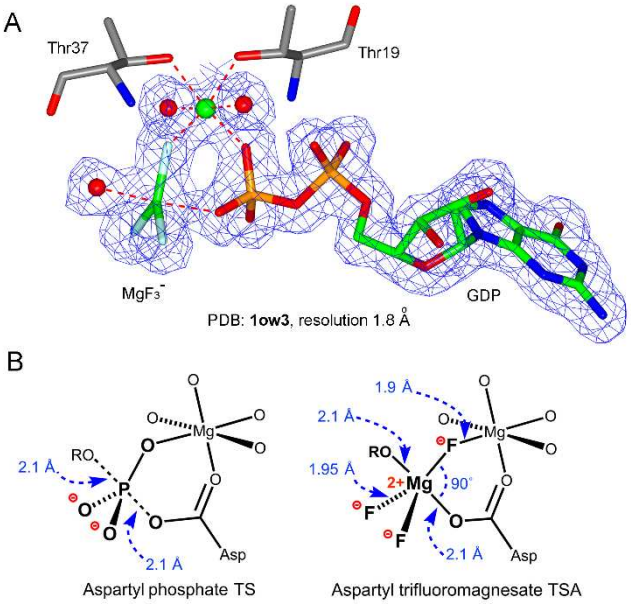


Figure. 11 (A) MgF_3^- complex with GDP for RhoA (PDB: **1ow3**) showing electron density. (B) Typical MgF_3^- complexes with aspartate residues in a six-membered ring complex with the catalytic Mg^{2+}.

4.2. Aluminum Trifluoride, AlF_3^0

The first example of an aluminum trifluoride complex was presented in 1997 for the tbp complex in the active site of a dinucleotide kinase (PDB: **1kdn**), shortly to be followed by a study on Ras/RasGAP with a GDP complex.[26] There are now 56 examples of structures that report an aluminum trifluoride core. Of these, three are octahedral (Section 3.2), and three have been shown by ^{19}F NMR to be trifluoromagnesates (see Sections 4.1 and 7.2) Of the remainder, only two alkaline phosphatase structures may be identified as having a tbp aluminum trifluoride core (Fig. 12). In mutant P300A (PDB: **1kh5**) two catalytic Zn^{2+} ions share one fluorine while Ser102 and a zinc-coordinated water provide the axial ligands for the tbp aluminum. It has an apical O–Al–O distance of 3.80 Å and Al-F bonds of 1.75 Å characteristic

of the AlF_4^- complexes described above (Section 3.1, SI Table 7). What is the situation for the remaining 48 AlF_3^0 complexes?

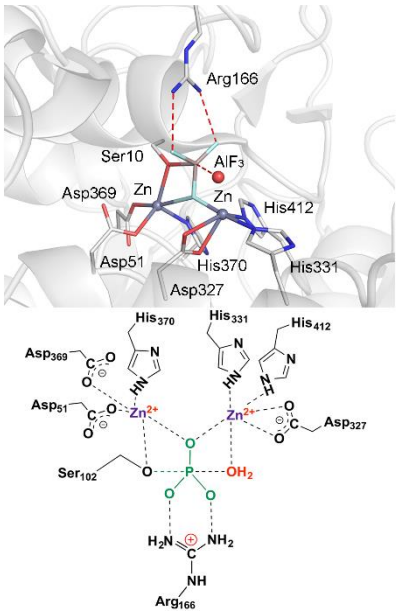


Figure. 12 Upper**:** Structure of the catalytic center for alkaline phosphatase complexed to AlF_3 (PDB: **1kh5**). Lower: Cartoon of the coordination organization in the active site with transferring phosphoryl group (green) and nucleophilic water (red).

The influence of pH on the transition between octahedral and tbp structures of AlF_x complexes in protein crystal structures for phosphoryl transfer enzymes was proposed to involve a switch from AlF_4^- to AlF_3^0 at elevated pH.[27] However, studies on the pH dependence of aluminum hydroxide solubility supported an alternative interpretation.[14] Aluminum hydroxide precipitates at pH ≥ 8, hence it will be replaced by magnesium in the protein complexes, with a consequent change to tbp geometry. That conclusion has now been validated by pH-dependent ^{19}F NMR analyses for several proteins (Section 7.2).[24b, 28] In some boundary cases, e.g. cAPK and PSP, there is partial dual occupancy of the active site by tbp and octahedral complexes in the crystal.[19, 24b, 24c] In structural terms, the dimensions of the tbp complexes closely reflect those of known trifluoromagnesates: axial O-M-O bonds 4.29 ± 0.39 Å, and M-F bonds 1.75 ± 0.12 Å. It is therefore likely that ^{19}F NMR analysis or crystallization in an aluminum-free medium will justify reassignment of some, or many, of these complexes as trifluoromagnesates (SI Table 9).

Taken together with trifluoromagnesates, a common general pattern of axial ligands emerges. The MF_3 species requires at least one anionic oxygen. ADP (25) and GDP (10) phosphates provide the overwhelming majority of examples while aspartate (11) is also significant. Water (27) is the dominant neutral axial ligand while serine and threonine hydroxyls appear infrequently. There is no example of both axial ligand positions occupied by two neutral ROH groups. As was observed for octahedral complexes (Section 3.1.4), there is only one example with histidine as a ligand (PDB: **1kdn**).(NB: Protein tyrosine phosphatases use a cysteine – histidine ion pair mechanism).[29]

4.3. Tetrafluoromagnesate, $MgF_4^=$

A group of structures for the Ca^{2+} pump ATPase contain tetrahedral moieties that have been assigned as $MgF_4^=$ without further experimental validation. Magnesium is only rarely 4-coordinate and then usually has sterically-bulky ether oxygens as ligands.[30] In all the examples in the PDB, the tetrahedral $MgF_4^=$ moiety is remote from ADP, coordinated to magnesium, with one or more of its atoms in contact with a backbone carbonyl oxygen (e.g. PDB: **1wpg**).[31] Subsequent work has described the same tetrahedral moiety for the Na/K pump ATPase (PDB: **2zxe**).[32] However, this "$MgF_4^=$" is proximate to a magnesium that has an aspartate ligand that closely resembles the 6-membered ring tbp structure common for complexes of aspartate with MgF_3^- (Section 4.1 and Fig. 15C). We therefore re-refined the structure, with results described below (Section 7.3). We conclude that a better chemical interpretation for all these "$MgF_4^=$" situations is that they are trifluoromagnesates that mimic the TS for hydrolysis of an aspartyl phosphate.

Finally, the most remarkable MF_x structure is that of a human diphosphoinositol phosphatase, cocrystallized with *myo*-inositol hexa*kis*-phosphate and then soaked with sodium fluoride (PDB: **2q9p**).[33] The resulting complex has four octahedral magnesiums with nine ligands assigned as fluorines. This complex embraces MgF_2, MgF_3, MgF_4, and MgF_5 species in a single complex and offers the first example of octahedral MgFx (Fig. 13). Its core appears related to the Rutile structure of MgF_2 that has octahedral magnesium and trigonal planar fluorine.[34]

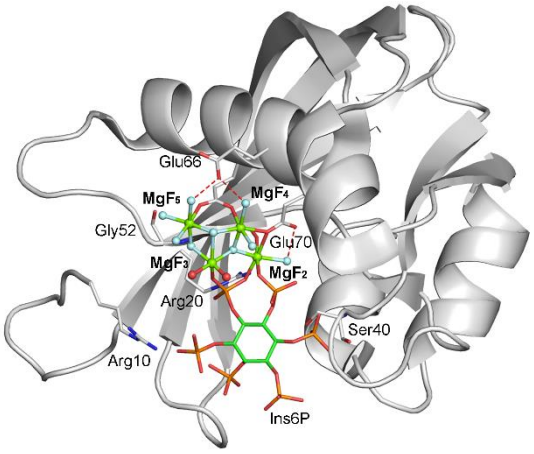


Figure. 13 Structure of hPPIP5K2 (PDB: **2q9p**) to show the "Mg_4F_9" cluster adjacent to phosphates 4 and 5 of Ins6P.

5. ^{19}F NMR Studies on MF_x

The inclusion of metal fluoride moieties within protein complexes has opened up the opportunity to use ^{19}F NMR measurements to examine the environment in which phosphate groups reside within the protein. The ^{19}F isotope has 100% natural abundance and a very high gyromagnetic ratio (25.18×10^7 $T^{-1}s^{-1}$), leading to very high sensitivity NMR spectra. Hence, metal fluoride species can be detected at low protein concentrations, and in large molecular weight complexes.[20, 24b, 24c, 35]

5.1 Chemical shifts

The chemical shifts of ^{19}F resonances provide a key measure of interactions between MF_x moieties and their protein hosts. They are reliable reporters of the electronic environment in the vicinity

of the fluorine nuclei and, by extrapolation and calculation, they act as reporters of the electronic environment experienced by the phosphoryl oxygen atoms in the TS.[20, 36] ^{19}F resonances display a high degree of dispersion and are predictable with good precision from quantum calculations of electronic distribution.[37] The average chemical shifts of resonances from AlF_x, MgF_x and BeF_x species differ (-154, -156, and -169 ppm, respectively), but a wide spread of individual shifts is observed in complexes with proteins. In similar βPGM complexes, for example, the average chemical shifts are -138 (AlF_4^-), -153 (MgF_3^-) and -160 (BeF_3^-) ppm.[13, 24c] This distribution is strongly affected by the vicinity of H-bond donors, as shown clearly in a comparison of the G6P and the 2-deoxyG6P complexes of βPGM. In the βPGM·MgF_3^-·2deoxyG6P TSA complex one fluorine loses its H-bond partner and its resonance moves substantially upfield (–18.1 ppm). (NB: ^{19}F chemical shifts are quoted relative to trifluoroacetic acid as reference]

The high sensitivity of ^{19}F chemical shifts to the surrounding environment can be used to show how enzymes control the influence of changes of protonation state. Thus, for βPGM it was observed that ^{19}F chemical shifts are invariant over the pH range 6.5 – 9.5, indicating that any changes in protonation state of the protein has no detectable influence on the environment of the TS complex. Characteristic average chemical shift values for different MF_x species have identified that millimolar fluoride is sufficiently effective at leaching Al^{3+} from glass, including borosilicate glass, to transform MgF_3^- complexes into AlF_4^- complexes unless an aluminum chelator such as deferoxamine is present.

5.2 Chemical exchange

It is observed, particularly in the $AlF_4^=$ complexes of some enzymes (including many early NMR studies of these complexes), that individual ^{19}F resonances coalesce to a single resonance as a result of rapid chemical exchange.[23b, 38] Resolved resonances of similar complexes have chemical shift differences of up to 10 kHz, and this shows that in some AlF_4^-complexes the interchange of fluorines greatly exceeds this rate. All MgF_3^-complexes of wild-type enzymes reported to date have resolved ^{19}F resonances, and hence much slower rates of fluorine interchange. For BeF_3^- complexes, the spectra show evidence of faster exchange rates than for MgF_3^- complexes.[13]

5.3 NOEs

Proton distribution in the vicinity of fluorine nuclei in the MF_x moiety can be assessed through the quantitation of ^{19}F–^{1}H NOEs. This approach has been used to determine solution structures of βPGM·MgF_3^-·G6P TSA and βPGM·AlF_4·G6P TSA complexes and so resolve a controversy concerning a reported pentaoxyphosphorane for this enzyme (Section 7.1).[24a, 24c] Traditionally ^{19}F–^{1}H NOEs are difficult to quantify owing to the effects of spin diffusion between ^{1}H nuclei as the ^{19}F–^{1}H NOE builds but, for MF_x complexes, the primary NOEs are to exchangeable protons. Hence ^{1}H-^{1}H spin diffusion can be suppressed by using a perdeuterated enzyme in a protonated buffer. Resonance assignment of the exchangeable ^{1}H nuclei in the protein allows unambiguous assignment of individual ^{19}F resonances.

5.4 SIIS – solvent induced isotope shifts for ^{19}F NMR

Proton distributions in the vicinity of fluorine nuclei can be assessed independently of ^{19}F–^{1}H NOEs on the basis of solvent induced hydrogen/deuterium primary isotope shifts (SIIS) of the ^{19}F resonances. For H-bonds to fluorine present between MF_x moieties and proteins, F····H–N and F····H–O, the magnitudes of the isotope shifts reflect local proton densities because of the through-space transmission of electric field differences between N–H and N–D bonds.[39] For example, in the βPGM·MgF_3^-·G6P TSA complex (Fig. 14), the three fluorines are coordinated by three protons (in a distorted tetrahedral arrangement), two protons (in a trigonal arrangement) and one proton, giving sum SIIS values of 1.6 ppm, 1.4 ppm, and 0.9 ppm, respectively. Comparing the G6P and the 2-deoxyG6P TSA complexes of βPGM, the sum SIIS value of one fluoride ion for the latter complex falls to close to zero (0.2 ppm), indicating that loss of the hexose 2-OH group leaves this fluorine virtually devoid of H-bonds.[28] The consequence of the removal of this hydroxyl group on the whole TSA complex is also observable as the other two fluorines move closer to their hydrogen bonding partners, as shown by small increases in their sum SIIS values (to 1.7 ppm and 1.5 ppm).

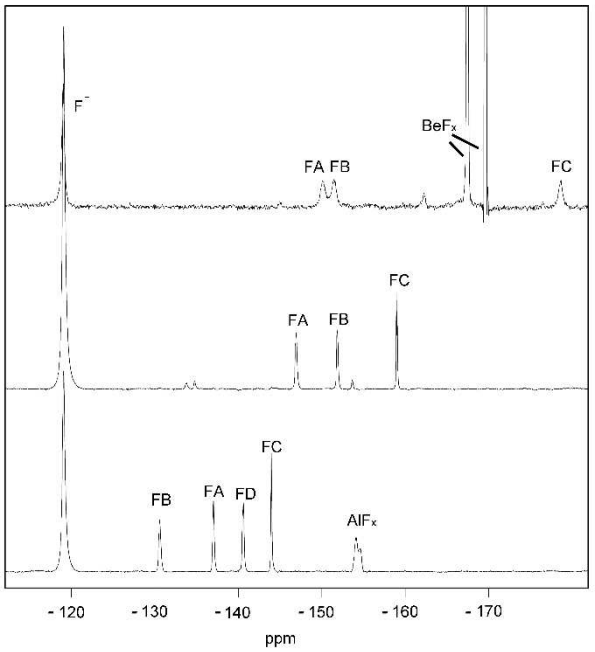


Figure. 14 1D ^{19}F NMR spectra of βPGM complexes with BeF_3^- (upper), MgF_3^- and G6P (middle), and AlF_4^- and G6P (lower). The ^{19}F resonance at -119 ppm in each spectrum is from free F^- ions, while those between -160 and -170 ppm (upper spectrum) are from unbound BeF_x species and those between -150 and -160 ppm (lower spectrum) are from unbound AlF_x species. The middle spectrum contains 3 small peaks from a second MgF_3^- bound protein conformation.

5.5 Scalar couplings across hydrogen bonds

Details of the coordination of the MF_x moiety by the protein is further shown in scalar couplings between nuclei involved with N-H····F H-bonds. $^1J_{HF}$ and $^2J_{NF}$ couplings have been reported for individual H^N····F pairs, with values up to 59 and 36 Hz, respectively.[36b] The magnitudes of both scalar couplings correlate closely with distances measured from crystal structure analysis. Hence, as well as reporting on the interaction across

individual hydrogen bonds, scalar couplings provide an independent means of assigning ^{19}F resonances, and cross-validating solution and crystal behavior.

5.6 Conclusions

NMR measurements of ^{19}F nuclei in the active site of MF_x TSA complexes provide a picture of the relationship between charge distribution of the mimic for phosphoryl group transfer and the enzyme. The good relationship between ^{19}F chemical shifts and SIIS values illustrates the dominant influence that the very local H-bonding groups have on shaping charge density on MF_x moieties. Moreover, the strong correlation between the observed NMR parameters and coordinates determined for numerous proteins in the crystalline state is a vital link showing that atomic positions determined at high resolution in the solid phase very closely reflect solution behavior.

6. Computational Analyses of MF_x Complexes

There have been relatively few computational studies of MF_x complexes within protein binding sites as such. Instead, these GSA and TSA structures have been widely used as starting points for a very large number of calculations by switching the MF_x atoms into PO3 while retaining the same geometry. Such computation has been applied to the molecular mechanisms of a variety of enzymes catalyzing phosphoryl transfer reactions,[40] particularly the small GTPases, which play critical roles in cell signaling and regulation, and to cAPK.[41] Theoretical methods provide considerable insights into the distribution of electrons within molecules, and the energies of protein/ligand interactions that mediate binding and TS stabilization.[42] In some cases, accurate structures can be obtained and used to resolve the nature of MF_x species in X-ray crystal structures of relatively low resolution.[43] Such calculations, briefly described below, yield useful information on the extent to which MF_x moieties resemble ground states or TSs in enzyme-catalyzed phosphoryl transfer.[20]

6.1 Computational Methods.

The principal approach to obtaining the properties of MF_x complexes has been the use of Density Functional Theory (DFT), given the ability of this method to yield accurate structural properties.[44] Numerous reviews are available that detail the theoretical principles underlying DFT together with its limitations, which include problems in modeling dispersion interactions and activation energy barriers in chemical reactions.[45] One important advantage of DFT, is that molecular systems composed of relatively large numbers of atoms can be treated completely quantum mechanically, allowing considerable insight into the electrostatic properties of MF_x complexes and how these might be perturbed by being in a protein environment. The general strategy has been to build active site models composed of the MF_x complex and residues that interact directly with the complex and surrounding molecules, such as ATP and GDP.[46] Larger models can also be built that include "second shell" residues, which form H-bonds to the initial set of inner residues.[20] In an alternative approach, which avoids the need to place artificial coordinate restraints on atoms in the QM region, the complete system is modeled using QM/MM methodologies.[47] Here the QM region is embedded in the rest of the protein and solvent, with the additional atoms (in an MM region) being described by classical potential energy functions that depend on "force field" parameters. Various methods can then be used to "couple" the QM and MM regions.[48] The advantage of the QM/MM approach, which also permits the inclusion of electrostatic effects arising from the protein and solvent environment, lies in the elimination of "edge effects" at the boundaries of the QM region arising from coordinate restraints. In addition, the relatively simple potentials used to describe the MM region allow the use of MD simulations to obtain free energy estimates for the system, which are not reliably obtained by analysis of the geometry-optimized QM active site models.[49]

6.2 BeF_3 complexes

As discussed in Section 2, beryllium fluoride complexes resemble GS phosphate groups when bound to nucleophilic groups or dinucleotides. The extent to which such tetrahedral complexes mimic phosphate moieties was explored using QM calculations of BeF_3^- complexed to the catalytically important aspartate side chain of βPGM in the presence and absence of βG6P, a substrate for the enzyme.[13] Large models, consisting of the BeF_3^- complex and 29 residues surrounding the active site, were obtained from crystal structures of these complexes and structurally optimized using B3LYP and 6-31G basis set, with the inclusion of d polarization functions for the fluoride ions.[13] As usual, the outer atoms in these models were constrained to their crystallographic coordinates. Atomic charges were then computed using the Mulliken formulation in order to minimize computational expense. The results showed that the beryllium and fluoride ions carry about 60% and 75% of the charges expected for phosphorus and oxygen atoms in a phosphate group. Hence, although the total charge of the BeF_3^- moiety is identical to that of the reactive intermediate in the enzyme-catalyzed reaction, the internal separation of charge is scaled down.[13]

6.3 MgF_3 complexes.

There is ample evidence that the MgF_3^- ion is an excellent stable analog of the TS for phosphate transfer in a number of enzyme-catalyzed reactions (Section 4). Early DFT calculations were performed to investigate the claim that X-ray crystallography had revealed the structure of a phosphorane intermediate in the reaction catalyzed by βPGM, and validated the correction that the tbp complex was MgF_3^- (Section 7.1).[50] The calculated distances for a MgF_3^- anion were consistent with those seen in the crystal structure. Subsequent high-level QM/MM calculations have supported this conclusion, and have shown that it also holds for phosphoryl transfer catalyzed by UTPase.[51] QM/MM studies followed that sought to demonstrate that MgF_3^- was present in medium-resolution X-ray crystal structures of the Ras/RasGAP complex rather than the isoelectronic AlF_3.[43] The QM region was modeled using standard Hartree-Fock *ab initio* calculations, which ignore the effects of electronic correlation. Nonetheless, this level of QM theory was sufficient to show that calculated distances and angles for the MgF_3^- complex were in much better agreement with the crystal structure for the Ras/RasGAP·GDP·MF_x complex than those computed for either AlF_3 or AlF_4^-. This was an important result because the electron density observed for the MF_x species in the Ras/RasGAP·GDP·MF_x structure (PDB: **1wq1**) was

inadequate to permit an unambiguous assignment of the ion.[26a] More recent work has sought to establish the extent to which MgF_3^- resembles the phosphoryl moiety undergoing transfer in the TS for GTP hydrolysis catalyzed by the RhoA·RhoGAP·GDP complex.[20] Specifically, this study, which employed DFT calculations on a very large active site model, demonstrated that the observed ^{19}F chemical shifts for the RhoA·RhoGAP·GDP·MgF_3^- complex can indeed be interpreted as indirect measures of the relative electron densities of the cognate oxygen atoms in the "true" TS for attack of water on the terminal phosphate of GTP.[20]

6.4 AlF_3 complexes.

Notwithstanding the questions raised about the validity of designating many tbp MF_x complexes as AlF_3, their structures, notably for Ras and for cAPK, have been used as starting points for many computations. The success of these computations lies in the simplicity of the transformation of AlF_3^0 into PO_3^- without regard to the change in charge involved. Only the tbp geometry matters.

6.5 Conclusions

Taken overall, the number of computational studies on the electronic structure and steric properties of protein-bound MF_x complexes has been relatively limited. There has also been limited evaluation of their resemblance to TS structures calculated using either QM or QM/MM methods for a range of enzymes, and their dynamic behavior within the active site remains poorly explored. This is surprising given the clear differences in the ^{19}F NMR spectra reported for BeF_3^-, MgF_3^- and AlF_4^- containing complexes (Section 5).

MF_x complexes have necessarily provided valuable starting points for numerous QM and QM/MM studies of the mechanism(s) of phosphoryl transfer, with particular focus on the Ras·RasGAP·GDP·MF_x structure (**1wq1**) as a basis for efforts to model the structure and energetics of the TS for Ras-catalyzed GTP hydrolysis.[26a] This choice has not, however, led to a consensus view of the mechanism. For example, extensive QM/MM calculations by some groups consistently predict a partially associative reaction on the basis of careful free energy estimates (Fig. 2).[40, 49, 52] On the other hand, other workers have reported a variety of QM and QM/MM studies in which they present evidence for a loose (dissociative) TS (Fig. 2).[43, 53] Similarly, there is substantial disagreement about the true functional role of a conserved active site glutamine, particularly regarding whether it mediates proton transfer.[40, 54] Finally, the number of waters that might participate in proton transfer has also been a subject of debate. In contrast to what is observed in MF_x complexes, it is argued that a critical proton transfer to substrate requires an second water molecule in addition to that which is the nucleophile in GTPase-catalyzed hydrolysis (Section 8.5).[55] The energetic penalty for introducing this "second" water is estimated, however, to be within thermal energy when PDB: **1wq1** is used as the initial model in QM/MM calculations.[56] While such disparate conclusions may reflect inherent differences in the computational methods chosen to model reaction mechanism, and the inclusion or absence of adequate conformational sampling, it is possible that the quality of the MF_x-containing crystal structure might influence the calculations, especially if extensive equilibration using dynamics is not performed prior to geometry optimization and TS location.[49] As we point out above, there is considerable variation in the quality of structures deposited in the PDB.

7. Sorting the Sheep and the Goats

Studies on MF_x transcend the boundary between protein crystallography and biomolecular chemistry. As a result, many situations exist which can benefit from closer integration of the available experimental and computational approaches. Several examples have been identified where electron density data can be reassigned by a broader approach to its interpretation, while this review has identified new examples capable of reanalysis. This is notably where the electron density maps are insufficiently resolved to make their interpretation unambiguous in the absence of a chemical evaluation. We briefly highlight two cases that are fully documented and one that is susceptible of reinterpretation.

7.1. MgF_3^- misidentified as PO_5.

The 2003 publication of a tbp complex in the active site of βPGM as a pentaoxyphosphorane received immediate attention, and re-examination.[24a, 28, 50a, 57] A combination of computation (Section 6.3) and ^{19}F NMR analysis (Section 5) established that it is accurately interpreted as a trifluoromagnesate complex (Fig. 15A).[28, 50b] A later in-depth QM/MM analysis calculated both the reaction path of the phosphorylation step (using PO_3^-) and also the geometry of a complex with the MgF_3^- TSA. It concluded that trifluoromagnesate is a good mimic of the true TS, which has pentacoordinate phosphorane character.[51a]

7.2. MgF_3^- misidentified as AlF_3.

An authoritative and extensive study on protein kinase A (cAPK) included the description of a tbp complex for the phosphorylation of a target serine peptide by ATP.[58] ^{19}F NMR established the major presence of MgF_3^- in the complex along with some octahedral AlF_4^-, showing that charge balance predominates over geometry in selection of the TS analog (Section 4.2, Fig. 13B).[24b, 36a] This result has been endorsed by DFT computation on cAPK (Fig.15B).[59]

7.3. MgF_3^- misidentified as $MgF_4^=$.

We have explained that it is exceptional to find magnesium in the form of tetrahedral tetrafluoromagnesate, $MgF_4^=$ (Section 4.2). Of 28 examples of this tetrahedral ligand listed in the PDB, the best resolved (2.40 Å, PDB: **2zxe**) is for an ATPase ion pump from shark. Its electron density map does not unambiguously support interpretation as a magnesium-coordinated tetrahedral $MgF_4^=$.[31-32] We have therefore re-refined the data to show an alternative interpretation of a tbp MgF_3^- covalently bonded to the essential Asp376 (Fig. 15C). This has an axial O-Mg-O distance of 3.85 Å, an in-line angle of 171.3°, and Mg-F bonds 1.86 Å. It is likely that this analysis could be applied to some or all of these tetrahedral complexes, although electron density is not available for the majority of them.

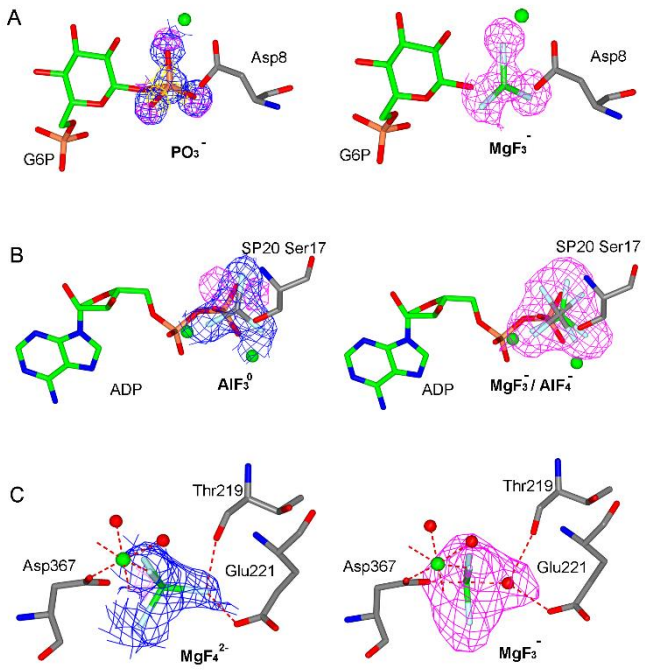


Figure. 15 The electron densities for the original interpretation and re-interpretation based on the unbiased omit map *F*o - *F*c. (A) The original map for pentaoxyphosphorane in PDB: **1o08** (left) and the unbiased omit map for MgF_3^- in PDB: **2wf5** (right). (B) The original map for AlF_3^0 in PDB: **1l3r** (left) and the unbiased omit map for the reinterpretation mixed occupancy for MgF_3^-/AlF_4^- at 70/30 ratio. (C) The original map for MgF_4^{2-} in PDB: **2zxe** (left) and the omit map (right) for an alternative MgF_3^- and water in the same density. All the unbiased *F*o - *F*c omit maps in magenta are contoured at 3σ for the metal fluoride moiety before their inclusion in the model and the 2*F*o-*F*c maps in blue are contoured at 1σ.

8. Fundamentals of Phosphoryl Transfer revealed by MF_x

8.1 Protein conformation – H-bonded and aligned NACs

The accessibility of high-resolution structures and solution NMR measurements for multiple MF_x complexes allows a detailed picture to be developed of many of the steps involved in catalysis. βPGM is a very good example where data are available for the apo-enzyme, the BeF_3^- mimic of the phosphoenzyme (EP), the BeF_3^- mimic of the EP complexes with both substrates (G6P and βG1P), and the corresponding MgF_3^- and AlF_4^- TSA complexes for each reaction. From them the development of the TS complex can be mapped out (Fig. 16). These data reveal how the EP down-regulates hydrolysis by disfavoring water from occupancy of a position to attack the phosphate. The EP undergoes domain closure in the presence of substrate but to alternative NACs.[13] The first is a more stable complex where the substrate H-bonds with the target phosphate, and which interconverts with a second, less stable complex where the substrate is aligned for attack. The latter NAC develops into the TS. This mutase operates on each of its two substrates in two consecutive reactions. A comparison of its behavior with the two substrates reveals that the protein conformation is conserved in the TSs of the two chemical steps, and the enzyme responds to the step change in substrate geometry by utilizing water molecules as spacers in one reaction, and leaving the transferring phosphate group depleted in H-bond partners in the other.[35b]

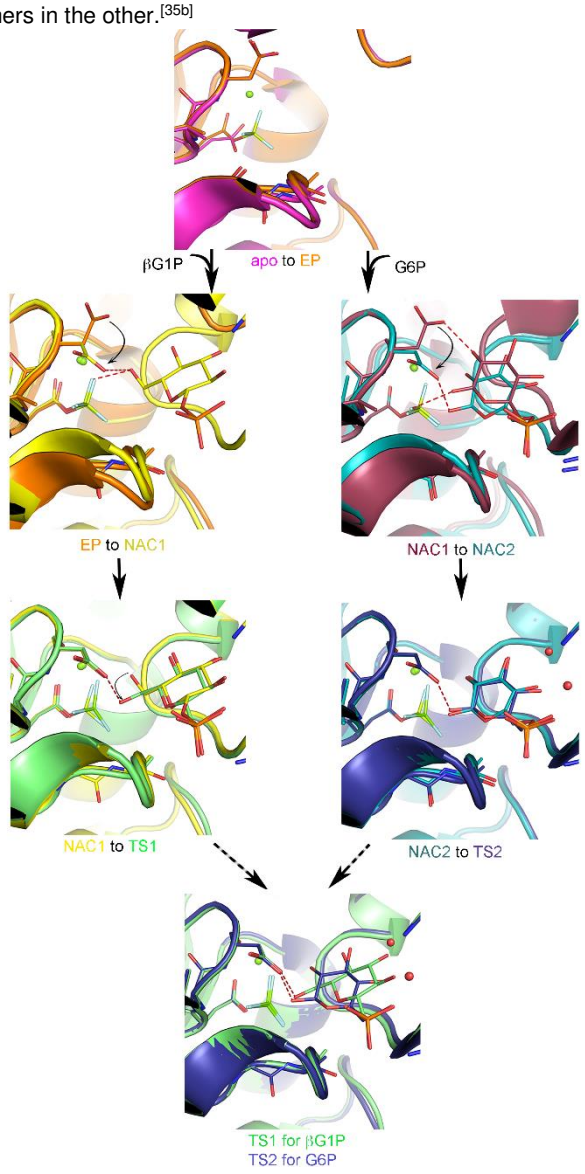


Figure. 16 Progression of βPGM active site from ground state (top, magenta) to transition state (bottom) (rainbow coloring shows pairwise progression). Left track: Step 1 pathway via phosphoenzyme (EP) (orange) to NAC1 (yellow) to TS1 for phosphorylation of glucose β-1-phosphate (green). Right track: Step 2 pathway via NAC1 (grape) to NAC2 (cyan) to TS2 (deep blue). Domain closure (EP to NAC) is linked to conformational adjustment of catalytic Asp10 to provide GABC for the glucose-OH group.

8.2 Charge Balance – Neutralize the "Anionic Shield"

The concept of charge balance was provoked by the observation that Ap_5A (5 –ve charges) is a better inhibitor of adenylate kinase than is Ap_4A (4 –ve charges).[10a] The true TS (6 –ve charges) is thus better mimicked by Ap_5A. The concept says that enzymes complement the excess anionic charge on TSs for phosphoryl transfer by cationic Mg^{2+} and side-chain residues in the immediate vicinity of the transferring phosphorus atom. Studies on hPGK validated this concept by demonstrating that hPGK prioritizes anionic charge over geometry in selection of MF_x for TSA complex

formation.[60] Based on the geometry of MF_x complexes for a wide range of phosphoryl transfer enzymes, it was demonstrated that charge balance is maintained within a sphere of up to 15 Å around the transferring phosphorus even when that borders on bulk water (Fig. 17).[60] A classic example is that of cAPK where charge balance is only achieved by the incursion of the substrate peptide with three +ve charges into a 13.5 Å sphere.[24b] This concept has been endorsed in a DFT study on cAPK, that found the order of affinity to the enzyme is MgF_3^-> AlF_4^- > AlF_3 while it confirmed charge balance out to 8 Å from the reaction center.[59]

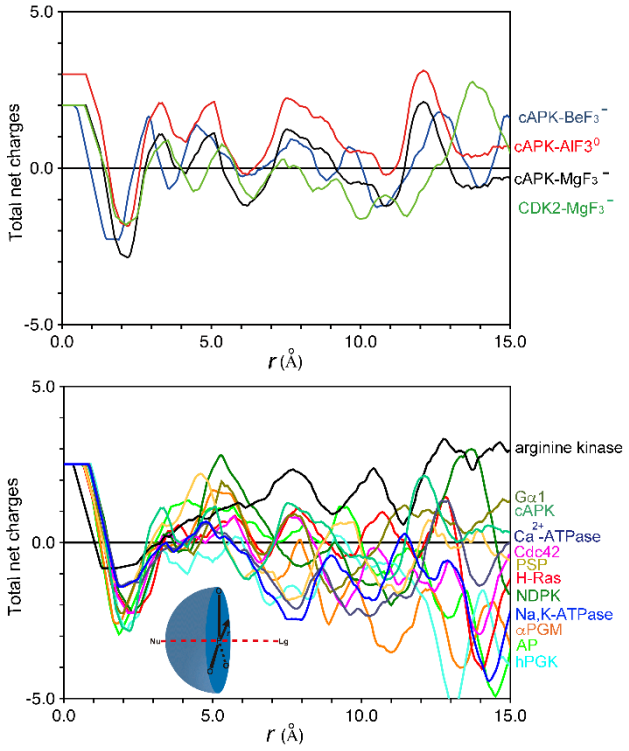


Figure. 17 Upper: Charge balance for kinases cAPK and CDK2 showing distortion for "AlF_3^0" assignment. Lower: Charge balance for a range of phosphoryl transfer proteins with insert showing radial nature of charge balance calculation.

8.3 Optimize Geometry – "In-Line" phosphoryl transfer

"In-Line" nucleophilic substitution at phosphorus was established in the 1980s by elegant stereochemical work, although in geometric terms it was a rather coarse measure.[61] Over a hundred MF_x structures have refined that analysis, with the 30 highest resolution AlF_4^- and MgF_3^- TSA complexes having "in-line" angles with a mean value of 175.2˚± 2.6˚. These same structures have revealed much more than the simple "in-line" geometry. A steadily growing number of examples in the PDB deliver reactant, TSA, and product structures for the same enzyme. In ten cases to date, they can be aligned not only to fine-tune "in-line" phosphoryl transfer but to provide a picture of the process at atomic resolution. The key chemistry takes place within a trigonal bipyramid whose apices are the donor (Od) and acceptor (Oa) oxygens and the three equatorial oxygens. In the TS, phosphorus (or its surrogate metal ion) lies in the medial plane, shifting 1.2 Å from its position in the donor complex in the reactant to its position in the acceptor complex for the product (Fig. 18). The equatorial oxygens have the same coordination to amino acids and catalytic metals in the three states and change position by less than 0.4 Å from reactant to product (Table 1). The distance between Od and Oa contracts in the progression from reactant to the TS by 0.5 Å and then expands by 0.3 Å in the product complexes. Overall, these data give validity to the concerted nature of phosphoryl transfer and establish that it is primarily a ***phosphorus transfer*** process!

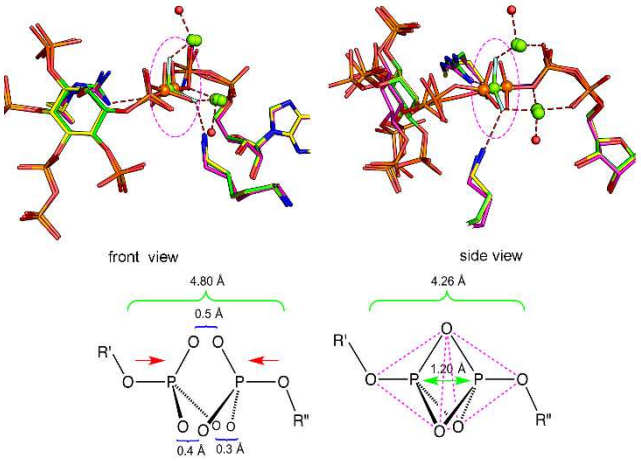


Figure. 18 Upper: Structures aligned (backbone Cα) for phosphoryl transfer by human hPPIP5K2. Reactants (red), transition state (yellow) and product (green) complexes show "in-line" transfer of phosphoryl group from ADP (right) to Ins6P (left) with near superposition of the three equatorial oxygens of the tbp in side and orthogonal front view. **Lower**: Approach of reactants by 0.4 Å places three equatorial oxygens in TS locations enabling phosphorus to move 1.2 Å through the core of the tbp complex to effect phosphoryl transfer.

8.4 Desolvation – Activate the nucleophile and the electrophile

The importance for catalysis of the exclusion of water from the active site of phosphoryl transfer enzymes historically has proponents[62] and opponents.[63] The overwhelming majority of x-ray structures in this review on MF_x as a TSA for phosphoryl transfer show that there are commonly only two situations where water comes within 4 Å of the transferring phosphorus. First, this is where the water is an isolated nucleophile for the hydrolysis of ATP, GTP or an aspartyl phosphate. Secondly, it is where water features as a ligand for coordination to a catalytic Mg^{2+} that is itself coordinated to the transferring phosphoryl group. Thus, for 10 well-resolved ADP·AlF_4^- complex structures, the average distance to the next nearest non-specific water is 4.3 ± 0.7 Å. It is also evident that water is more clearly excluded from the catalytic center in MF_x structures of TSA complexes than in the corresponding NAC structures. Thus for 12 small G proteins the next nearest water is 6.6 ± 0.2 Å for GDP·AlF_4^- TSAs but 4.22 ± 0.1Å for NACs. The unavoidable conclusion is that model studies on the hydrolysis of ATP and GTP in water cannot reliably be extrapolated to enzyme catalysis. A prime reason for exclusion of water is the control of H-bonding to neutral OH nucleophiles. Without exception all of these show proximity to a H-bond acceptor, often an aspartate carboxylate.[10b] While this interaction has historically been interpreted as providing GABC, recent computational analyses have shown that proton transfer occurs

late in the TS, as especially demonstrated for the small G protein, RhoA (Section 8.5).[20, 50b, 64]

Such H-bonding activity is equally evident in PGM, PSP, and phosphoglycerate mutase. Its purpose is primarily to orientate the oxygen for nucleophilic attack by enabling orbital overlap and denying H-bonding from the OH group to the anionic oxygens of the electrophilic phosphoryl group. Strong support is provided from a study on RNase A in which His12 and His119 were independently replaced by 4-fluorohistidine, pK_a 3.5. The artificial mutants delivered an unchanged k_{cat} but with greatly modified pH profiles.[65] This result is consistent with these histidines delivering H-bonding for nucleophile orientation rather than for GABC. While our interpretation is founded on MF_x complexes in proteins, it may well also apply to enzyme mechanisms for C-O cleavage, typically glycosidases, which have hitherto been designated as using GABC.

8.5 GTP hydrolysis depends on controlling H-bonds

Small G proteins accelerate the hydrolysis of bound GTP to GDP by 10^{11} using a mechanism whose details have been very controversial.[20, 40] In particular, LFER and KIE studies have supported a proposal that the hydrolysis of GTP in water is a dissociative process.[66] This analysis has been extrapolated to the Ras-catalyzed reaction,[67] and recently endorsed by KIE analysis as having a loose TS.[68] Aqueous hydrolysis has been developed in a QM study that has invoked a second water molecule to assist in proton transfer in the TS.[52, 56] This proposition has been developed into a "two water" mechanism for enzymatic hydrolysis of GTP based on a modest resolution structure for Ras at 2.5 Å resolution (PDB: **1wq1**) which has a less closed assembly of residues in the TS.[69]

What is the evidence from MF_x studies? To date, over 30 octahedral and tbp x-ray structures of GDP·MF_x TSA complexes can be superposed to show that water attacks Pγ "in-line" (Fig. 19A) in trigonal coordination with H-bonds donated to Thr37 and Gln63 (RhoA numbering), and in a compact TS.[20] Moreover, there is no second water in any of the high-resolution TSA structures, the next nearest water being 6 Å distant from Pγ (excepting two waters coordinating the catalytic Mg^{2+}). The ^{19}F NMR spectrum of a RhoA/RhoGAP·GDP·MgF_3^- TSA complex has identified F1 as the most shielded fluorine and DFT computation extends that analysis to O1G as the most electronegative oxygen. High-level QM calculations, using 87 heavy atoms drawn from 17 amino acids, show that, for RhoA/RhoGAP, the MgF_3^- complex accurately mimics the true TS for phosphoryl transfer. It involves neither torsional phosphate strain nor GABC, and has an "in-line" angle of 175° with an O–P–O distance of 4.27 Å in a tight TS. The primary barrier to GTP hydrolysis appears to be the propensity of water to H-bond to an oxygen on the terminal phosphoryl group, as shown for 18 structures of small G proteins with GPPNP that have the water H-bonded to O2G (Fig. 19B). This denies orbital overlap between nucleophile and electrophile. Thus, the core of the catalytic mechanism is the orientation of both protons on the water *away from GTP* by passive H-bonds. It enables the nucleophilic oxygen to effect occupied orbital overlap with the antibonding orbital of Pγ (Fig. 19C). GABC is not needed in GTPases, as has been confirmed by computation showing that the protons remain on oxygen in the TS.[20]

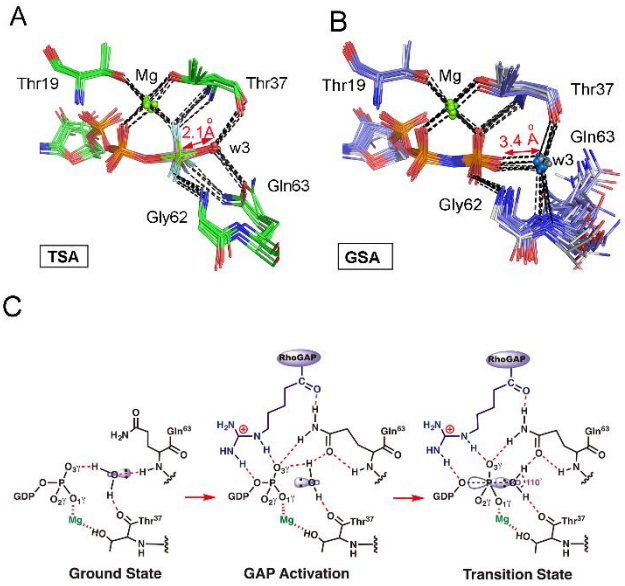


Figure. 19 (A) Catalytic site for 8 small G proteins in tbp GDP.MF_x complexes (green). Nucleophilic water complexed to M (2.1 Å) in-line and H-bonded to Thr37 and Gln63. **(**B) Catalytic site for GSA structures of 18 small G proteins with GPPNP (blue) H-bonded to water at 3.4 Å separation in NAC complexes.[20] (C) Cartoon showing change in water orientation from GS to intermediate stage and to TS through completion of the H-bond network by GAP protein. [20]

9. Conclusions

The three primary MF_x species are trifluoroberyllate, tetrafluoroaluminate, and trifluoromagnesate. Structural, spectroscopic, and computational methods have combined to validate their use as surrogates for the phosphoryl group in ground state and transition state analog complexes for a wide variety of enzymes. The results achieved through their use have delivered details of phosphoryl transfer at the atomic level and supported investigations of protein folding and aggregation for tertiary structure problems. However, their use has been overwhelmingly committed to studies on terminal, dianionic phosphates and their reactions, with hardly any incursion into phosphate diester chemistry. That remains a major challenge for the future.

Acknowledgements

The authors thank Drs. Nicola Baxter, Matthew Bowler, Matthew Cliff and Robert Molt for valuable discussions and input during the preparation of this review. This work was supported by the BBSRC and University of Sheffield, UK, and by Cardiff University, UK. Y. J. is funded by ERC Advanced Grant AdG-322942.

Keywords: Metal fluorides • transition state analogs • phosphoryl transfer • enzyme mechanisms • ^{19}F NMR spectroscopy • DFT analysis

References:

[1] a) A. S. Mildvan, *Proteins* **1997**, *29*, 401-416; b) W. W. Cleland, A. C. Hengge, *Chem. Rev.* **2006**, *106*, 3252-3278; c) J. K. Lassila, J. G. Zalatan, D. Herschlag, *Annu. Rev. Biochem.* **2011**, *80*, 669-702.
[2] R. E. Mesmer, C. F. Baes, *Inorg. Chem.* **1969**, *8*, 618-626.
[3] a) A. J. Fisher, C. A. Smith, J. B. Thoden, R. Smith, K. Sutoh, H. M. Holden, I. Rayment, *Biochemistry* **1995**, *34*, 8960-8972; b) G. D. Henry, S. Maruta, M. Ikebe, B. D. Sykes, *Biochemistry* **1993**, *32*, 10451-10456; c) J. P. Issartel, A. Dupuis, C. Morat, J. L. Girardet, *Eur. Biophys. J.* **1991**, *20*, 115-126.
[4] G. M. Blackburn, J. Cherfils, G. P. R. Moss, N.G. J. Nigel, J. P. Waltho, N. H. Williams; A. Wittinghofer, *Pure Appl. Chem.* **2016**, *in press*.
[5] L. Pauling, *The nature of the chemical bond, Vol. E3*, Cornell University Press, New York, **1960**.
[6] a) S. Hur, T. C. Bruice, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 12015-12020; b) T. C. Bruice, *Chem. Rev.* **2006**, *106*, 3119-3139.
[7] S. S. Batsanov, *Inorg. Mater.* **2001**, *37*, 871-885.
[8] E. S. Burgos, M. C. Ho, S. C. Almo, V. L. Schramm, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 13748-13753.
[9] I. Schlichting, J. Reinstein, *Biochemistry* **1997**, *36*, 9290-9296.
[10] a) G. E. Lienhard, I. I. Secemski, *J. Biol. Chem.* **1973**, *248*, 1121-1123; b) M. W. Bowler, M. J. Cliff, J. P. Waltho, G. M. Blackburn, *New J. Chem.* **2010**, *34*, 784-789.
[11] G. M. Blackburn, *Chem. Ind. (London)* **1981**, *7*, 134–138.
[12] a) E. Kowalinski, A. Schuller, R. Green, E. Conti, *Structure* **2015**, *23*, 1336-1343; b) A. K. Park, J. H. Lee, Y. M. Chi, H. Park, *Biochem. Biophys. Res. Commun.* **2016**, *473*, 625-629; c) S. R. Sheftic, E. White, D. J. Gage, A. T. Alexandrescu, *Biochemistry* **2014**, *53*, 311-322; d) S. Maruta, Y. Uyehara, T. Aihara, E. Katayama, *J. Bio. chem.* **2004**, *136*, 57-64; e) B. J. Hilbert, J. A. Hayes, N. P. Stone, C. M. Duffy, B. Sankaran, B. A. Kelch, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E3792-E3799; f) O. Pylypenko, W. Attanda, C. Gauquelin, M. Lahmani, D. Coulibaly, B. Baron, S. Hoos, M. A. Titus, P. England, A. M. Houdusse, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 20443-20448.
[13] J. L. Griffin, M. W. Bowler, N. J. Baxter, K. N. Leigh, H. R. W. Dannatt, A. M. Hounslow, G. M. Blackburn, C. E. Webster, M. J. Cliff, J. P. Waltho, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 6910-6915.
[14] a) R. Bruce Martin, *Biochem. Biophys. Res. Commun.* **1988**, *155*, 1194-1200; b) R. Bruce Martin, *Coord. Chem. Rev.* **1996**, *149*, 23-32.
[15] P. C. Sternweis, A. G. Gilman, *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 4888-4891.
[16] J. Bigay, P. Deterre, C. Pfister, M. Chabre, *EMBO J.* **1987**, *6*, 2907-2913.
[17] T. Higashijima, M. P. Graziano, H. Suga, M. Kainosho, A. G. Gilman, *J. Biol. Chem.* **1991**, *266*, 3396-3401.
[18] a) J. Sondek, D. G. Lambright, J. P. Noel, H. E. Hamm, P. B. Sigler, *Nature* **1994**, *372*, 276-279; b) D. E. Coleman, A. M. Berghuis, E. Lee, M. E. Linder, A. G. Gilman, S. R. Sprang, *Science* **1994**, *265*, 1405-1412.
[19] W. Wang, H. S. Cho, R. Kim, J. Jancarik, H. Yokota, H. H. Nguyen, I. V. Grigoriev, D. E. Wemmer, S. H. Kim, *J. Mol. Biol.* **2002**, *319*, 421-431.
[20] Y. Jin, R. W. Molt, J. P. Waltho, N. G. J. Richards, G. M. Blackburn, *Angew. Chem. Int. Ed.* **2016**, *55*, 3318-3322.
[21] Glyn R. Hemsworth, D. González-Pacanowska, Keith S. Wilson, *Biochem. J.* **2013**, *456*, 81-88.
[22] Y. Fovet, J.-Y. Gal, *Talanta* **2000**, *53*, 617-626.
[23] a) D. L. Graham, P. N. Lowe, G. W. Grime, M. Marsh, K. Rittinger, S. J. Smerdon, S. J. Gamblin, J. F. Eccleston, *Chem. Biol.* **2002**, *9*, 375-381; b) D. L. Graham, J. F. Eccleston, C. W. Chung, P. N. Lowe, *Biochemistry* **1999**, *38*, 14981-14987.
[24] a) N. J. Baxter, L. F. Olguin, M. Golicnik, G. Feng, A. M. Hounslow, W. Bermel, G. M. Blackburn, F. Hollfelder, J. P. Waltho, N. H. Williams, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 14732-14737; b) Y. Jin, M. J. Cliff, N. J. Baxter, H. R. W. Dannatt, A. M. Hounslow, M. W. Bowler, G. M. Blackburn, J. P. Waltho, *Angew. Chem. Int. Ed.* **2012**, *51*, 12242-12245; c) N. J. Baxter, G. M. Blackburn, J. P. Marston, A. M. Hounslow, M. J. Cliff, W. Bermel, N. H. Williams, F. Hollfelder, D. E. Wemmer, J. P. Waltho, *J. Am. Chem. Soc.* **2008**, *130*, 3952-3958.
[25] B. U. Klink, R. S. Goody, A. J. Scheidig, *Biophys. J.* **2006**, *91*, 981-992.
[26] a) K. Scheffzek, M. R. Ahmadian, W. Kabsch, L. Wiesmuller, A. Lautwein, F. Schmitz, A. Wittinghofer, *Science* **1997**, *277*, 333-338; b) K. Rittinger, P. A. Walker, J. F. Eccleston, S. J. Smerdon, S. J. Gamblin, *Nature* **1997**, *389*, 758-762.
[27] I. Schlichting, J. Reinstein, *Nat. Struct. Biol.* **1999**, *6*, 721-723.
[28] N. J. Baxter, A. M. Hounslow, M. W. Bowler, N. H. Williams, G. M. Blackburn, J. P. Waltho, *J. Am. Chem. Soc.* **2009**, *131*, 16334-16335.
[29] Z. Y. Zhang, J. E. Dixon, *Biochemistry* **1993**, *32*, 9340-9345.
[30] C. W. Bock, A. Kaufman, J. P. Glusker, *Inorg. Chem.* **1994**, *33*, 419-427.
[31] C. Toyoshima, H. Nomura, T. Tsuda, *Nature* **2004**, *432*, 361-368.
[32] T. Shinoda, H. Ogawa, F. Cornelius, C. Toyoshima, *Nature* **2009**, *459*, 446-450.
[33] A.G. Thorsell, C. Persson, S. Gräslund, M. Hammarström, R. D. Busam, B. M. Hallberg, *Proteins: Struct., Funct., Bioinf.* **2009**, *77*, 242-246.
[34] W. H. Baur, *Acta Cryst.* **1956**, *9*, 515-520.
[35] a) L. Xiaoxia, J. P. Marston, N. J. Baxter, A. M. Hounslow, Z. Yufen, G. M. Blackburn, M. J. Cliff, J. P. Waltho, *J. Am. Chem. Soc.* **2011**, *133*, 3989-3994; b) Y. Jin, D. Bhattasali, E. Pellegrini, S. M. Forget, N. J. Baxter, M. J. Cliff, M. W. Bowler, D. L. Jakeman, G. M. Blackburn, J. P. Waltho, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 12384-12389.
[36] a) K. N. Leigh, C. E. Webster, *Dalton Trans.* **2014**, *43*, 3039-3043; b) N. J. Baxter, M. W. Bowler, T. Alizadeh, M. J. Cliff, A. M. Hounslow, B. Wu, D. B. Berkowitz, N. H. Williams, G. M. Blackburn, J. P. Waltho, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 4555-4560.
[37] E. Oldfield, *Philos. Trans. R. Soc. Lond. **B** Bio.l Sci.* **2005**, *360*, 1347-1361.
[38] a) G. R. Hoffman, N. Nassar, R. E. Oswald, R. A. Cerione, *J. Biol. Chem.* **1998**, *273*, 4392-4399; b) G. J. Praefcke, M. Geyer, M. Schwemmle, H. Robert Kalbitzer, C. Herrmann, *J. Mol. Biol.* **1999**, *292*, 321-332.
[39] J. G. Sośnicki, M. Langaard, P. E. Hansen, *J. Org. Chem.* **2007**, *72*, 4108-4116.
[40] S. C. Kamerlin, P. K. Sharma, R. B. Prasad, A. Warshel, *Q. Rev. Biophys.* **2013**, *46*, 1-132.
[41] A. Pérez-Gallegos, M. Garcia-Viloca, À. González-Lafont, J. M. Lluch, *ACS Catalysis* **2015**, *5*, 4897-4912.
[42] E. Brunk, U. Rothlisberger, *Chemical Reviews* **2015**, *115*, 6217-6263.
[43] B. L. Grigorenko, A. V. Nemukhin, R. E. Cachau, I. A. Topol, S. K. Burt, *J. Mol. Model.* **2005**, *11*, 503-508.
[44] P. Geerlings, F. De Proft, W. Langenaeker, *Chem. Rev.* **2003**, *103*, 1793-1874.
[45] a) A. J. Boone, C. H. Chang, S. N. Greene, T. Herz, N. G. J. Richards, *Coord. Chem. Rev.* **2003**, *238–239*, 291-314; b) J. Antony, S. Grimme, *Phys. Chem.Chem.Phys.* **2006**, *8*, 5287-5293; c) P. E. M. Siegbahn, *J.Biol. Inorg. Chem.* **2006**, *11*, 695-701.
[46] F. Himo, *Theor. Chem. Acc.* **2006**, *116*, 232-240.
[47] H. M. Senn, W. Thiel, *Angew. Chem. Int. Ed.* **2009**, *48*, 1198-1229.
[48] M. J. Field, P. A. Bash, M. Karplus, *J. Comput. Chem.* **1990**, *11*, 700-733.
[49] M. Klahn, S. Braun-Sand, E. Rosta, A. Warshel, *J. Phys. Chem. B* **2005**, *109*, 15645-15650.
[50] a) S. D. Lahiri, G. Zhang, D. Dunaway-Mariano, K. N. Allen, *Science* **2003**, *299*, 2067-2071; b) C. E. Webster, *J. Am. Chem. Soc.* **2004**, *126*, 6840-6841.
[51] a) E. Marcos, M. J. Field, R. Crehuet, *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 2405-2411; b) I. Berente, T. Beke, G. Náray-Szabó, *Theor. Chem. Acc.* **2007**, *118*, 129-134.
[52] B. R. Prasad, N. V. Plotnikov, A. Warshel, *J. Phys. Chem. B* **2013**, *117*, 153-163.
[53] a) B. L. Grigorenko, A. V. Nemukhin, M. S. Shadrina, I. A. Topol, S. K. Burt, *Proteins: Struct., Funct., Bioinf.* **2007**, *66*, 456-466; b) I. A. Topol, R. E. Cachau, A. V. Nemukhin, B. L. Grigorenko, S. K. Burt, *Biochim. Biophys. Acta* **2004**, *1700*, 125-136.
[54] a) M. G. Khrenova, B. L. Grigorenko, A. B. Kolomeisky, A. V. Nemukhin, *J. Phys. Chem. B* **2015**, *119*, 12838–12845; b) A. Cavalli, P. Carloni, *J. Am. Chem. Soc.* **2002**, *124*, 3763-3768.
[55] A. Shurki, A. Warshel, *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 1-10.
[56] B. R. Prasad, N. V. Plotnikov, A. Warshel, *J. Phys. Chem. B* **2013**, *117*, 153-163.
[57] G. M. Blackburn, N. H. Williams, S. J. Gamblin, S. J. Smerdon, *Science* **2003**, *301*, 1184.
[58] Madhusudan, P. Akamine, N. H. Xuong, S. S. Taylor, *Nat. Struct. Biol.* **2002**, *9*, 273-277.
[59] A. J. M. Ribeiro, M. J. Ramos, P. A. Fernandes, N. Russo, *Chem. Phys. Lett.* **2013**, *571*, 66-70.
[60] M. J. Cliff, M. W. Bowler, A. Varga, J. P. Marston, J. Szabo, A. M. Hounslow, N. J. Baxter, G. M. Blackburn, M. Vas, J. P. Waltho, *J. Am. Chem. Soc.* **2010**, *132*, 6507-6516.
[61] a) J. R. Knowles, *Annu. Rev. Biochem.* **1980**, *49*, 877-919; b) P. A. Frey, *Adv. Enzymol. Relat. Areas Mol. Biol* **1989**, *62*, 119-201; c) G. Lowe, *Acc. Chem. Res.* **1983**, *16*, 244-251.
[62] V. E. Anderson, M. W. Ruszczycky, M. E. Harris, *Chem. Rev.* **2006**, *106*, 3236-3251.
[63] a) A. Warshel, J. Aqvist, S. Creighton, *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5820-5824; b) R. Wolfenden, *Biophys. Chem.* **2003**, *105*, 559-572.
[64] a) A. J. Smith, Y. Li, K. N. Houk, *Org. Biomol.Chem.* **2009**, *7*, 2716-2724; b) M. Valiev, R. Kawai, J. A. Adams, J. H. Weare, *J. Am. Chem. Soc.* **2003**, *125*, 9926-9927; c) A. Perez-Gallegos, M. Garcia-Viloca, A. Gonzalez-Lafont, J. M. Lluch, *Phys. Chem. Chem. Phys.* **2015**, *17*, 3497-3511.
[65] D. Jackson, J. Burnier, C. Quan, M. Stanley, J. Tom, J. Wells, *Science* **1994**, *266*, 243-247.
[66] S. J. Admiraal, D. Herschlag, *Chemical Biology* **1995**, *2*, 729-739.
[67] K. A. Maegley, S. J. Admiraal, D. Herschlag, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 8160-8166.
[68] X. Du, S. R. Sprang, *Biochem.* **2009**, *48*, 4538-4547.
[69] R. P. B, N. V. Plotnikov, J. Lameira, A. Warshel, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 20509-20514.

Table 1. PDB Triple Structure Overlays for Ten Proteins

Protein	PO_3^- Donor	PO_3^- Acceptor	PDB1 Reactant complex	PDB2 TSA complex	PDB3 Product complex	$P_r \cdots P_p$ dist Å	$O_r \cdot\cdot O_p$ dist Å OG1[d]	$O_r \cdot\cdot O_p$ dist Å OG2[d]	$O_r \cdot\cdot O_p$ dist Å OG3[d]	$O_r \cdot\cdot O_p$ dist Å global	Od····Oa distance reactant	Od····Oa distance TSA	Od····Oa distance product	Od·····Oa distance global	O–P–O angle TSA
*eco***Acid Pase**	AspP	Water	**2heg**	**2hf7**	**1rmy**	1.43	0.48	0.54	0.45	0.49	5.0	4.21	4.50	4.57	170.23
AK	ATP	AMP	**1ank**	**3sr0**	**4cf7**	1.24	0.59	1.00	0.66	0.75	4.53	4.17	4.71	4.47	173.20
cAPK	ATP	SerOH	**1rdq**	**1l3r**[a]	**1rdq**	1.06	-0.50	0.51	0.26	0.09	4.52	4.28	4.33	4.30	162.18
hPGK	ATP	3PGA	**4axx**	**2wzb**	**2x15**	1.21	0.23	0.58	0.59	0.15	4.55	4.27	4.54	4.58	170.91˚
βPGM	AspP	G1P	*tbp*	**2wf5**	**2wf8**	1.30	0.55	0.58	0.22	0.45	n/a	4.20	4.41	4.30	176.45
hPPIP5K2	ATP	InsP7	**3t9c**	**3t9e**	**3t9f**	1.36	0.40	0.50	0.58	0.49	4.66	4.20	4.66	4.84	167.13˚
PSP	AspP	SerOH	**1l7p**	**1l7n** [a]	**1j97**	0.98	0.18	-0.48	0.28	0.00	5.07	4.24	*5.45*	4.79	173.93
Rab11a	GTP	Water	**1oiw**	**1grn**	**1oix**	1.10	0.43	-0.48	0.76	0.24	n/a	4.39	4.68	4.55	*157.49*˚
Ras	GTP	Water	**1ctq**	**1wq1**	**1xd2**	1.39	0.65	0.81	*1.15*[c]	0.73	*6.22*[c]	4.45	4.67	4.61	165.13˚
RhoA.GAP	GTP	Water	**1a2b**	**1ow3**	**5fme**[b]	0.93	-0.66	0.38	0.53	0.08	5.24	4.19	4.44	4.62	172.38
Mean ± SD						**1.20 ± 0.18**	**0.24 ± 0.46**	**0.39 ± 0.49**	**0.48 ± 0.19**	**0.37 ± 0.41**	**4.80 ± 0.30**	**4.26 ± 0.09**	**4.55 ± 0.14**	**4.65 ± 0.51**	**170.2˚ ± 4.6˚**

(a) Rerefined (by Dr Matt Bowler) as MgF_3^- on the basis of ^{19}F NMR analysis (c) Unreleased

(b) Data in italics is ≥ 2 S.D. from the mean and so omitted from analysis (d) Order for the three O---O distances is clockwise (with catMg behind) and O1G coordinated to magnesium[4]